

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 90.3624 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVYTIQNGKMSSTIVSE.....ATVLDKNNISSKSTNNPNK 144

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	748	100.0	773	4	AAB48343
2	748	100.0	2120	3	AAY81710
3	748	100.0	2140	6	ABU01020
4	748	100.0	2140	6	ABU45746
5	748	100.0	2140	8	ADM92113
6	748	100.0	2140	8	ADT50099
7	745	99.6	637	8	ADR94534
8	745	99.6	637	9	AEA58404
9	745	99.6	2138	8	ADK48759
10	615	82.2	117	2	AAW55096
11	615	82.2	117	5	ABF54590
12	615	82.2	117	7	ADC45149
13	110	14.7	746	4	AAG81779
14	110	14.7	778	5	ABP39023
15	110	14.7	778	8	ADS06368
16	108	14.4	188	9	ADZ79639
17	108	14.4	354	9	ADZ72933
18	106.5	14.2	470	8	ADT56185
19	106.5	14.2	484	3	AAG47777
20	105	14.0	647	9	ADZ79635
21	105	14.0	651	8	AD019012
22	105	14.0	651	8	AD019010
23	104	13.9	665	3	AB18278
24	104	13.9	665	7	ABO23606

25	103.5	13.8	169	9	ADZ79634	P. falcip
26	102.5	13.7	775	6	ABU42797	Protein e
27	101.5	13.6	564	4	ABB61977	Abu42797
28	99.5	13.3	707	6	ABU25018	Abu25018
29	96	12.8	639	9	ADW88472	Staphyloc
30	95.5	12.8	635	8	ADS93954	AdS93954
31	95.5	12.8	635	8	ADV83292	AdV83292
32	95.5	12.8	643	8	ADV89902	AdV89902
33	95.5	12.8	643	8	ADV81155	AdV81155
34	95	12.7	645	9	ADW88460	Staphyloc
35	95	12.7	645	9	ADW88459	Staphyloc
36	95	12.7	645	9	ADW88458	Staphyloc
37	94	12.6	645	9	ADW88454	Staphyloc
38	94	12.6	645	9	ADW88457	Staphyloc
39	94	12.6	645	9	ADW88453	Staphyloc
40	94	12.6	645	9	ADW88441	Staphyloc
41	94	12.6	645	9	ADW88455	Staphyloc
42	94	12.6	645	9	ADW88456	Staphyloc
43	94	12.6	645	9	ADW88452	Staphyloc
44	93	12.4	645	9	ADW88451	Staphyloc
45	93	12.4	1791	8	ADP25441	Plasmodiu

#### ALIGNMENTS

##### RESULT 1

ID	AAB48343	standard; protein; 773 AA.
XX	AAB48343;	
AC	AAB48343;	
DT	20-APR-2001	(first entry)
XX		
DE	S. pneumoniae	Spl30 polypeptide.
XX		
KW	Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;	
KW	bronchial; lung; blood; infection; immune response; immunotherapy;	
KW	antibacterial; auditory; vaccine.	
XX		
OS	Streptococcus pneumoniae.	
XX		
FN	WO200076540-A2.	
XX		
PD	21-DEC-2000.	
XX		
PF	09-JUN-2000; 2000WO-US015925.	
XX		
PR	10-JUN-1999; 99US-0138453P.	
XX		
PA	(MEDI-) MED IMMUNE INC.	
XX		
PI	Adamou JE, Choi GH;	
XX		
DR	WPI; 2001-112197/12.	
XX		
DR	N-PSDB; AAC84742.	
XX		
PT	New vaccines comprising Spi28 or Spi30 polypeptides, for treating and	
PT	preventing pneumococcal infections, particularly infections caused by	
PT	Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or	
PT	blood infections.	
XX		
PS	Claim 8; Page 51-54; 54pp; English.	
XX		
CC	The invention relates to novel immunogenic polypeptides, Spi28 and Spi30	
CC	from S. pneumoniae. Vaccines comprising the polypeptides are useful for	
CC	the treatment and prevention of pneumococcal infections, particularly	
CC	infections caused by Streptococcus, such as otitis media, nasopharyngeal,	
CC	bronchial, lung or blood infections. The antigens are used as immunogenic	
CC	agents to stimulate an immune response. The antisera and antibodies may	
CC	also be used in diagnosing and treating pneumococcal infections.	
CC	Recombinant polypeptides serve as a mechanism for stimulating production	
CC	of antibodies for use in passive immunotherapy, diagnostic reagents, and	

CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the S. pneumoniae Sp130 polypeptide  
 XX  
 SQ Sequence 773 AA;

Query Match 100.0%; Score 748; DB 4; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 7e-67;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSK 60  
 DB 630 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSK 689  
 QY 61 DTFIKPVFKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 690 DTFIKPVFKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 749  
 QY 121 KDVTATVLDKNNISSKSTTNNPNK 144  
 DB 750 KDVTATVLDKNNISSKSTTNNPNK 773

RESULT 2  
 AAY81710  
 ID AAY81710 standard; protein; 2120 AA.  
 AC AAY81710;  
 XX  
 DT 02-JUN-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae protein sequence ID3.  
 XX  
 KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 FN WO200006738-A2.  
 XX  
 PD 10-FEB-2000.  
 XX  
 PF 27-JUL-1999; 99WO-GB002452.  
 XX  
 PR 27-JUL-1998; 98GB-00016336.  
 PR 19-MAR-1999; 99US-0125329P.  
 XX  
 PA (MICR-) MICROBIAL TECHNIQS LTD.  
 XX  
 PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
 XX  
 DR WPI; 2000-195301/17.  
 DR N-PSDB; AA291806.  
 XX  
 PT Streptococcal proteins and polynucleotides useful for diagnosis,  
 PT treatment and prophylaxis of bacterial infections.  
 XX  
 PS Claim 2; Page 41-42; 76pp; English.  
 XX  
 CC This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
 CC infection. As the sequences can be used to treat S. pneumoniae infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
 CC meningitis  
 XX  
 SQ Sequence 2120 AA;

Query Match 100.0%; Score 748; DB 3; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-56;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSK 60  
 DB 1943 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSK 2002  
 QY 61 DTFIKPVFKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 2003 DTFIKPVFKIEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 2062  
 QY 121 KDVTATVLDKNNISSKSTTNNPNK 144  
 DB 2063 KDVTATVLDKNNISSKSTTNNPNK 2086

RESULT 3  
 ABU01020  
 ID ABU01020 standard; protein; 2140 AA.  
 AC ABU01020;  
 XX  
 DT 23-OCT-2003 (revised)  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain protein from coding region #590.  
 XX  
 KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
 KW gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae; type 4 strain.  
 XX  
 FN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB002163.  
 XX  
 PR 27-MAR-2001; 2001GB-00007658.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Maignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 DR N-PSDB; ASX06302.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 XX  
 PS Claim 1; SEQ ID NO 1180; 56pp; English.  
 XX  
 CC The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC ABS56454. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target



CC sequence contained within a Streptococcus nucleic acid sequence, where  
CC the first primer is substantially complementary to the target sequence  
CC and the second primer is substantially complementary to the complement of  
CC the target sequence, and where the parts of the primers having  
CC substantial complementarity define the termini of the target sequence to  
CC be amplified, assay comprising contacting a test compound with the  
CC protein, and determining whether the test compound binds to the protein  
CC and a Streptococcus pneumoniae bacterium, where one or more genes  
CC encoding the proteins has been rendered inactive. The proteins, nucleic  
CC acid molecules, antibody and compositions are useful as medicaments for  
CC treating or preventing a disease or infection due to streptococcus  
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
CC media or ear infection. They are also useful in developing vaccines,  
CC diagnostics and antibiotics. The methods are useful for identifying  
CC immunodominant proteins. The present sequence is one of the 2469 proteins  
CC expressed by the identified coding regions from the genomic sequence.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 2140 AA;

Query Match 100.0%; Score 748; DB 6; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.9e-66;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HRVTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLK 60  
DB 1963 HRVTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLK 2022  
QY 61 DTFIKVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 120  
DB 2023 DTFIKVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 2082  
QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
DB 2083 KDVTATVLDKNNISSKSTNNPNK 2106

RESULT 4  
ABU45746  
ID ABU45746 standard; protein; 2140 AA.  
XX  
AC ABU45746;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #31273.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA49616.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 73670; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2140 AA;

Query Match 100.0%; Score 748; DB 6; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.9e-66;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HRVTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLK 60  
DB 1963 HRVTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLK 2022  
QY 61 DTFIKVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 120  
DB 2023 DTFIKVPKKEEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 2082  
QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
DB 2083 KDVTATVLDKNNISSKSTNNPNK 2106

RESULT 5  
ADM92113  
ID ADM92113 standard; protein; 2140 AA.  
XX  
AC ADM92113;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
DE S pneumoniae antigenic protein sequence SeqID310.  
XX  
KW antibacterial; gene therapy; Streptococcus pneumoniae infection;  
KW antigenic.  
XX  
OS Streptococcus pneumoniae.  
XX

```
PN WO2004020609-A2.
XX 11-MAR-2004.
XX
XX 02-SEP-2003; 2003WO-US027401.
XX
XX 30-AUG-2002; 2002US-0407082P.
XX
XX (TUFT ) UNIV TUFTS.
XX
XX Camilli A, Hava DL;
PI WPI; 2004-239189/22.
XX N-PSDB; ADM91876.
XX
XX New Streptococcus pneumoniae nucleic acid molecules, useful for
PT diagnosing, treating and preventing active infections of Streptococcus
PT pneumoniae.
XX
XX Claim 27; SEQ ID NO 310; 123pp; English.
XX
XX This invention relates to novel isolated Streptococcus pneumoniae nucleic
XX acid molecules and the antigenic polypeptides encoded by them. The
XX invention may be useful for the production of compounds with an
XX antibacterial activity or for gene therapy. The nucleic acid molecules,
XX compositions and methods disclosed are useful for treating Streptococcus
XX pneumoniae infection. The present sequence is that of an S pneumoniae
XX protein of the invention.
XX
XX Sequence 2140 AA;
SQ
    Query Match          100.0%; Score 748; DB 8; Length 2140;
    Best Local Similarity 100.0%; Pred. No. 2.9e-66;
    Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKGADAGYVINLSK 60
DB 1963 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKGADAGYVINLSK 2022
QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSQKSDST 120
DB 2023 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSQKSDST 2082
QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
DB 2083 KDVTATVLDKNNISSKSTTNNPNK 2106
RESULT 6
ID ADT50099 standard; protein; 2140 AA.
XX
XX ADT50099;
XX
XX 13-JAN-2005 (first entry)
XX
XX S_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.
XX
XX hyperimmune serum reactive antigen; antibacterial; vaccine;
XX bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;
XX sepsis; meningitis.
XX
XX Streptococcus pneumoniae TIGR4.
XX
XX WO2004092209-A2.
XX
XX 28-OCT-2004.
XX
XX 15-APR-2004; 2004WO-EP003984.
XX
XX 15-APR-2003; 2003EP-00450087.
XX
XX (INTE-) INTERCELL AG.
XX
XX Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;
XX WPI; 2004-758335/74.
XX N-PSDB; ADT49955.
XX
XX New hyperimmune serum reactive antigens from Streptococcus pneumoniae,
XX and encoding nucleic acid molecules, useful for diagnosing, preventing or
XX treating S. pneumoniae infections.
XX
XX Disclosure; SEQ ID NO 177; 191pp; English.
XX
XX This invention relates to novel nucleic acids encoding hyperimmune serum
XX reactive antigens, or fragments derived thereof. Specifically, it refers
XX to antigens selected from peptides and serum reactive epitopes that can
XX be used in pharmaceutical compositions that exhibit antibacterial
XX activity. The present invention describes a composition (including the
XX nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)
XX that is useful for manufacturing a medicament such as a vaccine, which
XX can be used to treat or prevent bacterial infections, particularly S.
XX pneumoniae infections that cause pharyngitis, otitis media, pneumonia,
XX bacteraemia sepsis and meningitis. The antigen or its fragment may also
XX be used for isolating, purifying and/ or identifying an interaction
XX partner of the hyperimmune serum reactive antigen, as well as for
XX manufacturing a functional nucleic acid selected from aptamers and
XX Spiegelmers or for manufacturing a functional ribonucleic acid selected
XX from ribozymes, antisense nucleic acids and siRNA. This polypeptide
XX sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen
XX of the invention.
XX
XX Sequence 2140 AA;
SQ
    Query Match          100.0%; Score 748; DB 8; Length 2140;
    Best Local Similarity 100.0%; Pred. No. 2.9e-66;
    Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKGADAGYVINLSK 60
DB 1963 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKGADAGYVINLSK 2022
QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSQKSDST 120
DB 2023 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHSQKSDST 2082
QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
DB 2083 KDVTATVLDKNNISSKSTTNNPNK 2106
RESULT 7
ID ADT94534 standard; protein; 637 AA.
XX
XX ADT94534;
XX
XX 16-DEC-2004 (first entry)
XX
XX Novel S. pneumoniae protein sequence, SEQ ID 3169.
XX
XX Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
XX bacterial infection.
XX
XX Streptococcus pneumoniae.
XX
XX US6800744-B1.
XX
XX 05-OCT-2004.
XX
XX 30-JUN-1998; 98US-00107433.
XX
XX 02-JUL-1997; 97US-0051553P.
XX
XX 12-MAY-1998; 98US-0085131P.
XX
```

PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2004-697205/68.  
DR N-PSDB; ADR91931.  
XX  
XX New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 3169; 151pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridizable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae protein sequences. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.  
XX  
XX Sequence 637 AA;  
XX  
XX Query Match 99.6%; Score 745; DB 8; Length 637;  
XX Best Local Similarity 99.3%; Pred. No. 1.1e-66;  
XX Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 HRVTVTIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWEISGFGKDGAGVYINLSK 60  
XX 460 HRVTVTIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWEISGFGKDGAGVYINLSK 519  
XX  
XX 61 DTFTKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDST 120  
XX 520 DTFTKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDST 579  
XX  
XX 121 KDVTATVLDKNNISSKSTNNPNK 144  
XX 580 KDVTATVLDKNNISSKSTNNPNK 603  
XX  
XX RESULT 8  
XX AEA58404  
XX ID AEA58404 standard; protein; 637 AA.  
XX  
XX AC AEA58404;  
XX  
XX 25-AUG-2005 (first entry)  
XX  
XX Streptococcus pneumoniae ORP amino acid sequence SEQ ID NO:3169.  
XX  
XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;  
XX vaccine.  
XX  
XX Streptococcus pneumoniae.  
XX  
XX OS  
XX  
XX FN US2005136404-A1.  
XX

PD 23-JUN-2005.  
XX  
XX 10-JUL-2003; 2003US-00617320.  
XX  
XX 02-JUL-1997; 97US-0051553P.  
XX 12-MAY-1998; 98US-0085131P.  
XX 30-JUN-1998; 98US-00107433.  
XX  
XX (DOUC/) DOUCETTE-STAMM L A.  
XX (BUSH/) BUSH D.  
XX  
XX Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2005-477576/48.  
XX N-PSDB; AEA55801.  
XX  
XX New isolated nucleic acid molecules and encoded polypeptides useful for  
PT diagnosing, preventing or treating bacterial infections, particularly  
PT Streptococcus pneumoniae infection.  
XX  
XX Claim 5; SEQ ID NO 3169; 144pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule for detecting,  
CC preventing or treating pathological conditions resulting from bacterial  
CC infection. The isolated nucleic acid comprises: (a) any of the 2603  
CC nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence  
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the  
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide  
CC sequence of at least 8 nucleotides in length, where the sequence is  
CC hybridizable to a nucleic acid having any of the nucleotide sequences in  
CC (a). Also described: (1) a recombinant expression vector comprising the  
CC above nucleic acid operably linked to a transcription regulatory element;  
CC (2) a cell comprising the recombinant expression vector; (3) producing an  
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence  
CC consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)  
CC treating a subject for S. pneumoniae infection; (6) a recombinant or its  
CC substantially pure preparation of an S. pneumoniae polypeptide or its  
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;  
CC (7) a vaccine composition for preventing or treating an S. pneumoniae  
CC infection, comprising an amount of the above nucleic acid or polypeptide;  
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;  
CC (9) a computer readable medium having recorded the nucleotide sequences  
CC of AEA55236 to AEA57838; (10) a computer based system for identifying  
CC fragments of the Streptococcus genome of commercial importance. The  
CC composition and methods are useful for diagnosing, preventing or treating  
CC bacterial infections, particularly S. pneumoniae infection. The present  
CC sequence represents a S. pneumoniae ORP amino acid sequence from the  
CC present invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from the USPTO web site.  
XX  
XX Sequence 637 AA;  
XX  
XX Query Match 99.6%; Score 745; DB 9; Length 637;  
XX Best Local Similarity 99.3%; Pred. No. 1.1e-66;  
XX Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 HRVTVTIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWEISGFGKDGAGVYINLSK 60  
XX 460 HRVTVTIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWEISGFGKDGAGVYINLSK 519  
XX  
XX 61 DTFTKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDST 120  
XX 520 DTFTKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDST 579  
XX  
XX 121 KDVTATVLDKNNISSKSTNNPNK 144  
XX 580 KDVTATVLDKNNISSKSTNNPNK 603  
XX  
XX RESULT 9  
XX ADK48759  
XX ID ADK48759 standard; protein; 2138 AA.

XX ADK48759;  
AC  
XX  
DT 20-MAY-2004 (first entry)  
DE Streptococcus pneumoniae protein, Seq ID No 5274.  
XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.  
XX Streptococcus pneumoniae.  
OS  
PN US6699703-B1.  
XX  
PD 02-MAR-2004.  
XX  
PF 26-MAY-2000; 2000US-00583110.  
XX  
PR 02-JUL-1997; 97US-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
PR 30-JUN-1998; 98US-00107433.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;  
XX  
DR WPI: 2004-212399/20.  
DR N-PSDB; ADK46098.  
XX  
XX New nucleic acid molecules and polypeptides useful for diagnosing,  
PT preventing and treating pathological conditions resulting from bacterial  
PT infection, e.g. Streptococcus pneumoniae infection, and in drug  
PT screening.  
XX  
PS Disclosure; SEQ ID NO 5274; 301pp; English.  
XX  
CC The invention relates to isolated Streptococcus pneumoniae nucleic acids  
CC and polypeptides. The nucleic acids and proteins are useful for  
CC diagnosing, preventing and treating pathological conditions resulting  
CC from bacterial infection, such as S. pneumoniae infection. These may also  
CC be used for drug screening procedures. The present sequence represents a  
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence  
CC data for this patent did not appear in the printed specification but was  
CC obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.  
XX  
SQ Sequence 2138 AA;  
Query Match 99.6%; Score 745; DB 8; Length 2138;  
Best Local Similarity 99.3%; Pred. No. 5.8e-66;  
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKADAGYVINLSK 60  
Db 1961 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKADAGYVINLSK 2020  
QY 61 DTFIKPVFKKIEKEEENKPTFDYSKKQNPQVNHSQLNESHKREDLQREHSQKSDST 120  
Db 2021 DTFIKPVFKKIEKEEENKPTFDYSKKQNPQVNHSQLNESHKREDLQREHSQKSDST 2080  
QY 121 KDVTATVLDKNNISSKSTTNPNK 144  
Db 2081 KDVTATVLDKNNISSKSTTNPNK 2104  
RESULT 10  
AAW55096  
ID AAW55096 standard; protein; 117 AA.  
XX  
AC AAW55096;  
XX  
DT 02-OCT-1998 (first entry)  
XX  
DE Streptococcus pneumoniae SP0043 protein.

XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
KW detection; pneumonia; otitis media; meningitis.  
XX Streptococcus pneumoniae.  
OS  
PN WO9818930-A2.  
XX  
PD 07-MAY-1998.  
XX  
PF 30-OCT-1997; 97WO-US019422.  
PR 31-OCT-1996; 96US-0029960P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;  
XX  
DR WPI: 1998-272224/24.  
DR N-PSDB; AAV27357.  
XX  
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae  
PT - or their epitope-containing fragments, useful in protective or  
PT therapeutic vaccines, and for diagnosis.  
XX  
PS Claim 11; Page 62; 118pp; English.  
XX  
CC The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose  
XX  
SQ Sequence 117 AA;  
Query Match 82.2%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 1.8e-54;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 28 YKGELEKGYQFDGWEISGFEKGKADAGYVINLSKDTFIKPVFKKIEKEEENKPTFDYSK 87  
Db 1 YKGELEKGYQFDGWEISGFEKGKADAGYVINLSKDTFIKPVFKKIEKEEENKPTFDYSK 60  
QY 88 KQNPQVNHSQLNESHKREDLQREHSQKSDSTKDVTATVLDKNNISSKSTTNPNK 144  
Db 61 KQNPQVNHSQLNESHKREDLQREHSQKSDSTKDVTATVLDKNNISSKSTTNPNK 117  
RESULT 11  
ABP54590  
ID ABP54590 standard; protein; 117 AA.  
XX  
AC ABP54590;  
XX  
DT 04-SEP-2002 (first entry)  
XX  
DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.  
XX  
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
KW antibacterial; Streptococcal infection; detection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US2002061545-A1.

```
XX 23-MAY-2002.
PD
XX
XX
PF 22-JAN-2001; 2001US-00765272.
XX
XX
PR 30-OCT-1997; 97US-00961083.
XX
XX (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
XX WPI; 2002-479261/51.
DR N-PSDB; ABQ84825.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
PT and for preventing or attenuating disease caused by Streptococcus
PT infection.
XX
XX Claim 11; Page 29; 70pp; English.
XX
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
CC pneumoniae antigens have antibacterial activity and can be used in
CC vaccines. The S. pneumoniae antigens can also be used to prevent or
CC attenuate a Streptococcal infection in an animal. The polynucleotides
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
CC of S. pneumoniae ORFs (open reading frames) which are used in an example
CC from the present invention
XX
XX Sequence 117 AA;
SQ
Query Match 82.2%; Score 615; DB 5; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.8e-54;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 YKGELEKGYQFDGWISGFEKGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 87
DB 1 YKGELEKGYQFDGWISGFEKGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 60
QY 88 KDNQVNHSQLNESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 144
DB 61 KDNQVNHSQLNESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 117
RESULT 12
ADC45149
ID ADC45149 standard; protein; 117 AA.
XX
XX ADC45149;
XX
XX 18-DEC-2003 (first entry)
XX
XX S. pneumoniae antigenic protein SP043.
XX
XX Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX
XX Streptococcus pneumoniae.
XX
XX US6573082-B1.
XX
XX 03-JUN-2003.
XX
XX 28-MAR-2000; 2000US-00536784.
XX
XX 31-OCT-1996; 96US-0029960P.
PR
```

```
PR 30-OCT-1997; 97US-00961083.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
XX WPI; 2003-764574/72.
DR N-PSDB; ADC45148.
XX
XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
PT useful for producing vaccines for prevention or attenuation of infection
PT by Streptococcus pneumoniae.
XX
XX Example 1; SEQ ID NO 68; 58pp; English.
XX
XX The invention relates to an isolated polynucleotide consisting of a
CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
CC antigens. Also included are making a recombinant vector by inserting the
CC nucleic acid into a vector, an isolated polynucleotide consisting of at
CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
CC acids are useful as DNA vaccine against Streptococcus pneumoniae
CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
CC antigen nucleic acids are useful as probes for use in diagnostic methods
CC for detecting S. pneumoniae gene expression. The present sequence
CC represents an S. pneumoniae antigenic protein.
XX
XX Sequence 117 AA;
SQ
Query Match 82.2%; Score 615; DB 7; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.8e-54;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 28 YKGELEKGYQFDGWISGFEKGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 87
DB 1 YKGELEKGYQFDGWISGFEKGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 60
QY 88 KDNQVNHSQLNESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 144
DB 61 KDNQVNHSQLNESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 117
RESULT 13
AAG81779
ID AAG81779 standard; protein; 746 AA.
XX
XX AAG81779;
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis open reading frame protein sequence SEQ ID NO:652.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
XX endocarditis.
XX
XX Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US030782.
XX
XX 09-NOV-1999; 99US-0164258P.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX
XX Kimmerly WJ;
XX
XX WPI; 2001-316495/33.
DR N-PSDB; AAH52629.
DR
```

XX Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
PT useful for vaccinating against infections, e.g. endocarditis.  
XX  
PS Claim 18; Page 208; 2189pp; English.  
XX  
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
(II), given in AG81454 to AG83120, from *Staphylococcus epidermidis*. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the *S.*  
CC *epidermidis* polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX  
SQ Sequence 746 AA;

Query Match 14.7%; Score 110; DB 4; Length 746;  
Best Local Similarity 27.0%; Pred. No. 0.062;  
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;  
QY 5 VTIONGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW---EISGPE-----G 48  
Db 596 ITINGKQIKQOQSVKSGTKVLPKHSKVMMLTDGELTMP-DMTGWTKEVDLAFEDLTJKLVS 654  
QY 49 KKDAGYVIN--LSKDTFIKPVFKKIEEKEENKPTFDVS---KKKNPQVNHSQLNES 102  
Db 655 TKGNGFVNTQNSISKQIIL-----NKKIEVLSAEDTDDQKTDDESDN 701  
QY 103 HRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141  
Db 702 KSKKQKADHDHNTSSSTKN-----DKSNADSKNDSDD 734

RESULT 14  
ABP39023  
ID ABP39023 standard; protein; 778 AA.  
XX  
AC ABP39023;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE *Staphylococcus epidermidis* ORF amino acid sequence SEQ ID NO:3868.  
XX  
KW *Staphylococcus epidermidis*; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.  
XX  
OS *Staphylococcus epidermidis*.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-00134001.  
XX  
PR 14-AUG-1997; 97US-0055779P.  
PR 08-NOV-1997; 97US-0064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.  
DR N-PSDB; ABN91568.  
XX  
PT Novel isolated nucleic acid encoding a *Staphylococcus epidermis*  
PT polypeptide, useful for diagnosing and treating bacterial infections.  
PS Disclosure; SEQ ID NO 3868; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences can  
CC also be used in the diagnosis and treatment of bacterial infections,  
CC particularly *S. epidermidis* infections. The sequences can be used to  
CC screen for compounds able to interfere with the *S. epidermidis* life cycle  
CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from the USPTO web site  
XX  
SQ Sequence 778 AA;  
Query Match 14.7%; Score 110; DB 5; Length 778;  
Best Local Similarity 27.0%; Pred. No. 0.066;  
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;  
QY 5 VTIONGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW---EISGPE-----G 48  
Db 628 ITINGKQIKQOQSVKSGTKVLPKHSKVMMLTDGELTMP-DMTGWTKEVDLAFEDLTJKLVS 686  
QY 49 KKDAGYVIN--LSKDTFIKPVFKKIEEKEENKPTFDVS---KKKNPQVNHSQLNES 102  
Db 687 TKGNGFVNTQNSISKQIIL-----NKKIEVLSAEDTDDQKTDDESDN 733  
QY 103 HRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141  
Db 734 KSKKQKADHDHNTSSSTKN-----DKSNADSKNDSDD 766  
RESULT 15  
ADS06368  
ID ADS06368 standard; protein; 778 AA.  
XX  
AC ADS06368;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE *Staphylococcus epidermis* polypeptide seqid 5663.  
XX  
KW antibacterial; vaccine; antisense therapy; *Staphylococcus epidermidis*;  
KW recombinant expression vector; infection; computer readable medium;  
KW computer based system.  
XX  
OS *Staphylococcus epidermidis*.  
XX  
PN US2004147734-A1.  
XX  
PD 29-JUL-2004.  
XX  
PF 01-DEC-2003; 2003US-00724972.  
XX  
PR 08-NOV-1997; 97US-0064964P.  
PR 13-AUG-1998; 98US-00134001.  
PR 29-NOV-1999; 99US-00450969.  
XX  
PA (DOUC/) DOUCETTE-STAMM L.  
PA (BUSH/) BUSH D.  
XX  
PI Doucette-Stamm L, Bush D;  
XX  
DR WPI; 2004-580138/56.  
DR N-PSDB; ADS02596.  
XX



PD 27-APR-2005.  
XX  
XX  
PF 24-OCT-2003; 2003EP-00292673.  
XX  
XX 24-OCT-2003; 2003EP-00292673.  
PR (INSP ) INST PASTEUR.  
XX  
XX PA PA  
XX  
XX PI Drulhe P;  
XX  
XX WPI; 2005-323987/34.  
DR N-PSDB; AD272252.  
DR  
XX Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
PT falciparum, which encode proteins useful for preparing vaccine  
PT compositions against malaria.  
XX  
XX Disclosure; SEQ ID NO 2; 137pp; English.  
XX  
XX The present invention relates to the protection against malaria. More  
CC particularly, the invention pertains to a family of MSP-3 (merozoite  
CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
CC falciparum, highly conserved in P. falciparum strains, simultaneously  
CC expressed in P. falciparum at the erythrocytic stages and encoding  
CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at  
CC their N-terminal extremity and which are located at the merozoites  
CC surface. The characterization of this gene family enables the definition  
CC of immunogenic and vaccine compositions against P. falciparum. The  
CC present sequence is the P. falciparum MSP-3-1 protein.  
XX  
XX Sequence 354 AA;  
SQ  
Query Match 14.4%; Score 108; DB 9; Length 354;  
Best Local Similarity 23.2%; Pred. No. 0.035;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
QY 7 IQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGF--EGKKDAG-----VVINLS 59  
DB 181 VLKAEAS-----YDYL-----GWFGGVPKKEENMLSHLYVSKD 221  
QY 60 KDTFKIPVKIKIEKKE-----ENKPTFDVSKKQNPQVNHSLQNSHRKE 106  
DB 222 KENISKENDVDLDEKEEAETEEBELEKNEEETESISEDEEBEKEEENDKK 281  
QY 107 DLQREHSQKSDTKDVTATVLDKNNISKSTNN 141  
DB 282 EQEKEQSNENNQKDMEA-----QNLISKNNNN 311  
RESULT 18  
ADT56185  
ID ADT56185 standard; protein; 470 AA.  
XX  
XX AC ADT56185;  
XX  
XX DT 13-JAN-2005 (first entry)  
XX  
XX DE Plant polypeptide, SEQ ID 6262.  
XX  
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX  
XX Viridiplantae.  
XX  
XX US2004216190-A1.  
XX  
XX PD 28-OCT-2004.  
XX  
XX 18-DEC-2003; 2003US-00739930.  
PF

XX 28-APR-2003; 2003US-00424599.  
PR 28-APR-2003; 2003US-00425115.  
XX  
XX (KOVA/) KOVALIC D K.  
XX Kovalic DK;  
XX  
XX WPI; 2004-757369/74.  
DR  
XX  
XX New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.  
XX  
XX Claim 2; SEQ ID NO 6262; 14pp; English.  
XX  
XX The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant protein  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
XX  
XX Sequence 470 AA;  
SQ

Query Match 14.2%; Score 106.5; DB 8; Length 470;  
Best Local Similarity 20.3%; Pred. No. 0.074;  
Matches 36; Conservative 31; Mismatches 65; Indels 45; Gaps 5;  
QY 1 HRVTVTIQNGKEMSGSTIVSEEDFILPVYKGLKGYQFDGWEISGFE-----GKKDA 52  
DB 84 NRVTDTVQNNNGESK-----YVQDLARRIRYDE-EATGSQAQRIDHPNQKNV 131  
QY 53 GYVINLSKDTPIKVPFKKIEEKEENKPTFDVSKKQDN-----PQ 93  
DB 132 GITEKAFENSPIETSHRVDDNKRINNQNKNFTAAKSSENAVSRVSFGADHKRAEVMGRPM 191  
QY 94 VNHSLQNE-----SHRKEDLQREHSQKSDTKDVTATVLDKNNISKSTNNPNK 144  
DB 192 ENRDQVRQTSAEKSHRKNVTKSEKPRDQGVKKTAKOKDRNKEKKEETESINK 249



RESULT 19  
AAG47777  
ID AAG47777 standard; protein; 484 AA.  
XX  
AC AAG47777;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridization assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135623P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144844P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145919P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147933P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.







QY 58 LSKDTFIKVPFKIKBEKBE-----EKNKTPD-----VSKKKNPQVNHSLNESH 103  
Db 175 -----VKBGVKELEKBEKBEKISDDHKVBEENKSDDHKVBENKSKSDH 228  
QY 104 RKEDLQR-BEHSOKSDSTKDTATVTLVDKNNISKSTNNPNK 144  
Db 229 KLEVKVBEHEBDEE-----DKBEKSENKNKDNK 261

RESULT 25  
ADZ79634  
ID ADZ79634 standard; protein; 169 AA.  
XX AC ADZ79634;  
XX DT 14-JUL-2005 (first entry)  
XX DE P. falciparum merozoite surface protein 3, amino acid residues 212-380.  
XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
XX KW immunotherapy; malaria; antimalarial; vaccine.  
XX OS Plasmodium falciparum.  
XX FH Key Location/Qualifiers  
XX FT Region 1..169  
XX FT /note= "Amino acid residues 212-380 of MSP3"  
XX PN WO2005040206-A1.  
XX PD 06-MAY-2005.  
XX PF 22-OCT-2004; 2004WO-BP012910.  
XX PR 24-OCT-2003; 2003US-00691672.  
XX PA (INSP ) INST PASTEUR.  
XX PI Drulhe P;  
XX DR WPI; 2005-355821/36.  
XX CHimeric molecule useful for preparing vaccine composition against  
PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
PT immunized with molecule.  
XX Claim 2; SEQ ID NO 2; 79pp; English.

The invention relates to a chimeric molecule that comprises a glutamate-rich protein (GLURP) moiety consisting of a polypeptide fragment (amino acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380 of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises antibodies against both polypeptides in mice immunized with it. Also described are: (i) a conjugate comprising the chimeric molecule of the invention bound to a solid support, (ii) an immunogenic composition comprising the chimeric molecule, the conjugate described above, or a mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine against malaria comprising the chimeric molecule, the conjugate described above, or a mixture of GLURP and MSP3 antigens as an immunogen, in association with a suitable vehicle, (iv) use of purified and/or recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a medicament against malaria, and (v) a medicament for passive immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP antibodies. The chimeric molecule of the invention or a mixture of GLURP and MSP3 antigens are useful for the preparation of a vaccine composition against malaria. This sequence represents Plasmodium falciparum MSP3 protein (amino acid residues 212-380).

XX Sequence 169 AA;  
SQ

Query Match

13.8%; Score 103.5; DB 9; Length 169;

Best Local Similarity 25.2%; Pred. No. 0.037;  
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;  
QY 11 KEMSTIVSEEDFILPVYKLEKGYQPDGWEISGF--EGKKDAG-----YVINLSKDTF 63  
Db 1 KEASS-----YDYL-----GWFGGVPEHKKEENMLSHLYVSSDKKENI 41  
QY 64 IKPVFKKIEKKBEENKPTPDVSKKKNPQVNHSLN-----ESHRKEDLQR 110  
Db 42 SKENDDVLDL-KBEAEABTEEELEEKNEETESISEDEBEEREEKEENKKEQEK 100  
QY 111 BEHSOKSDSTKDTATVTLVDKNNISKSTNN 141  
Db 101 EQSNENNDDKKDMEA-----QNLISKNNNN 126

RESULT 26  
ABU42797  
ID ABU42797 standard; protein; 775 AA.  
XX AC ABU42797;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #28324.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Staphylococcus epidermidis.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX DR N-PSDB; ACA46667.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 70721; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 775 AA;

Query Match 13.7%; Score 102.5; DB 6; Length 775;  
Best Local Similarity 24.7%; Pred. No. 0.38;  
Matches 38; Conservative 26; Mismatches 61; Indels 29; Gaps 7;  
QY 5 VTIQKEMSTIVSEEDFILPVYK-----GELEKGYQFDGW-----EISGPF-----G 48  
Db 625 ITIGNGKQIKQOSVKSGTKVLPKSHKVMMLTDGELTMP-DMTGWTKEDEVLPEDLTKIKVS 683  
QY 49 KKDAGYVIN--LSKDTFFIKPVFK-----KIEEKKEENKPTFDVSKKKNPQVNHQS 98  
Db 684 TKGNGFVTTQSIKQGIKKNDKIEVSLSAETDDQDKTDESDSKKSKKDKVDNNSN 743  
QY 99 LNESHRKEDLQREHSQKSDSTKDTATVLDKNN 132  
Db 744 ASSSSKNEKSNADSKNDSDDSTNETSGS--ERNN 775

RESULT 27  
ID ABB61977  
AC ABB61977;  
XX ABB61977;  
DT 26-MAR-2002 (first entry)  
DE Drosophila melanogaster polypeptide SEQ ID NO 12723.  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
OS Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL06080.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
PT interactions.  
XX Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 564 AA;

Query Match 13.6%; Score 101.5; DB 4; Length 564;  
Best Local Similarity 24.5%; Pred. No. 0.31;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
QY 21 EDFILPVYKGELEKGYQFDGW-----EISGFEKKGAGYVI-----NLSKDTFIK 65  
Db 78 EDLDTPLSESRFSK--VFDGWNVDHDEHDHGVQEPSEALDDHDEHDDHDEDEDE 135  
QY 66 PVFKKIEKKEENKPT-----PDVSKKKNPQVNHSQLNESHRKEDLQREHSQKSDS 119  
Db 136 PLTEELSEELSEEEPEETDEDEPAADNEYEDEENNA--GENITAEAESEEEENDD 193  
QY 120 TKDVTATVLDKNNISSKST 138  
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 28  
ID ABB25018  
AC ABB25018;  
XX ABB25018;  
DT 19-JUN-2003 (first entry)  
DE Protein encoded by prokaryotic essential gene #10545.  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX Clostridium difficile.  
XX WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA28888.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 25; SEQ ID NO 52942; 1766pp; English.  
PS The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense







DE Streptococcus agalactiae protein sequence, SEQ ID 2296.  
 XX Antibacterial; Vaccine; bacterial infection.  
 XX Streptococcus agalactiae.  
 XX FR2824074-A1.  
 XX PD 31-OCT-2002.  
 XX PF 26-APR-2001; 2001FR-00005642.  
 XX PR 26-APR-2001; 2001FR-00005642.  
 XX PA (INSP ) INST PASTEUR.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Glaser P, Ruaniok C, Chevalier P, Frangeul L, Lalioui L;  
 XX PI Zouine M, Couve B, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst P;  
 XX DR WPI; 2004-101891/11.  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX Claim 6; SEQ ID NO 2296; 2687pp; French.  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
 CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
 CC agalactiae involved in the synthesis of amino acids, cell membranes,  
 CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
 CC phospholipid metabolism, nucleotide metabolism including purines,  
 CC pyrimidines and/or nucleosides, regulatory functions, replication,  
 CC transcription, translation, protein transport, adaptation to atypical  
 CC conditions, sensitivity to medicines and/or analogues, functions related  
 CC to transporters, biosynthesis of cofactors, prosthetic groups and  
 CC transporters, cell membrane proteins and cellular machinery. (I) are  
 CC useful for the detection and/or amplification of nucleic acids.  
 CC Pharmaceutical composition comprising (I) or (II) are useful for  
 CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
 CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
 CC contains 6617 sequence whereas the present patent only contains 2344  
 CC sequences.  
 XX SQ Sequence 643 AA;  
 Query Match 12.8%; Score 95.5; DB 8; Length 643;  
 Best Local Similarity 24.5%; Pred. No. 1.5;  
 Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;  
 QY 9 NGKMSSTIVSEEDFILP--VYKGL-----EKGYQFD-----GWEISGPEGKDGVI 56  
 Db 321 NTEPLTSLYLNKKEKFLPNIPYKNGKILREEDKYSFDEDEEFGNLLSYNKLKNEVL 380  
 QY 57 NLSKDTPIKPVFKKLEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS-- 114  
 Db 381 NITSTILKP-----PEQKIVED---FNPYSNLDNLEIKKIRLNGSQKQVEKTSPT 433  
 QY 115 QKSDTKDVTATVLDKN--NISKSTT 139  
 Db 434 PQKETVKEQTEQKVSQNTQVEKKSET 460  
 RESULT 33  
 ID ADV81155 standard; protein; 643 AA.  
 XX AC ADV81155;  
 XX 24-FEB-2005 (first entry)  
 XX

DE Streptococcus agalactiae protein, SEQ ID 2296.  
 XX Antibacterial; vaccine; bacterial infection.  
 XX Streptococcus agalactiae.  
 XX WO200292818-A2.  
 XX PD 21-NOV-2002.  
 XX PF 26-APR-2002; 2002WO-IB003059.  
 XX PR 26-APR-2001; 2001FR-00005642.  
 XX PA (INSP ) INST PASTEUR.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Glaser P, Ruaniok C, Chevalier P, Frangeul L, Lalioui L;  
 XX PI Zouine M, Couve B, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst P;  
 XX DR WPI; 2004-101891/11.  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX Claim 6; SEQ ID NO 2296; 439pp; French.  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transporters, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.  
 XX SQ Sequence 643 AA;  
 Query Match 12.8%; Score 95.5; DB 8; Length 643;  
 Best Local Similarity 24.5%; Pred. No. 1.5;  
 Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;  
 QY 9 NGKMSSTIVSEEDFILP--VYKGL-----EKGYQFD-----GWEISGPEGKDGVI 56  
 Db 321 NTEPLTSLYLNKKEKFLPNIPYKNGKILREEDKYSFDEDEEFGNLLSYNKLKNEVL 380  
 QY 57 NLSKDTPIKPVFKKLEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS-- 114  
 Db 381 NITSTILKP-----PEQKIVED---FNPYSNLDNLEIKKIRLNGSQKQVEKTSPT 433  
 QY 115 QKSDTKDVTATVLDKN--NISKSTT 139  
 Db 434 PQKETVKEQTEQKVSQNTQVEKKSET 460  
 RESULT 34  
 ID ADW88460 standard; protein; 645 AA.  
 XX AC ADW88460;  
 XX 21-APR-2005 (first entry)  
 XX







QY 79 NKPTPDV-----SKKNDPVOVNSOLNESHKEDIQ---REHSQKSDSTKDVLT-ATVL 128  
Db 470 NSAKKEATPATPKTPSPVEKESQKQDSQKQDNKQLPSVEKENDASSRGKDTATPK 529  
QY 129 DKNINSSKSTNNPNK 144  
Db 530 TKGEVSSSTT--PTK 543

**RESULT 40**

ADW88441  
ID ADW88441 standard; protein; 645 AA.

AC ADW88441:

DT 21-APR-2005 (first entry)

DE Staphylococcus aureus hybrid polypeptide 0657nHybrid3.

ORF0657n; vaccine; antibacterial; protein engineering;

KW Staphylococcus aureus infection: mutuin.

OS Staphylococcus aureus.

OS Synthetic.

PN WO2005009378-A2.

PD 03-FEB-2005.

22-JUL-2004; 2004WO-US023522.

24-JUL-2003: 2003US-0489840P.

PA (MERI ) MERCK & CO INC.

XX PI Anderson AS, Kuklin N, Jansen KU:

XX  
DR WPI: 2005-123069/13.

Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
PT useful for inducing protective immune response in humans against  
PT *Staphylococcus aureus* infection.  
PT

PS Claim 7; SEQ ID NO 10; 84pp; English.

The present sequence is that of a *Staphylococcus aureus* protein ORP0657n hybrid polypeptide denoted 0657nHybrid3. This is an example of claimed hybrid polypeptide immunogens of the invention ADW84439-ADW84474 that comprise a modified *S. aureus* ORP0657n sequence ADW88433-ADW88438 containing amino acid substitutions that increase sequence similarity to ORF0190 ADW88432. The hybrid polypeptides contain one or more epitopes for ORP0657n and ORF0190. They were designed by taking into account the similarity and differences between native ORP0657n and ORF0190 protein sequences. The invention also provides nucleic acids encoding these hybrid polypeptides, and a method for evaluating the ability of an immunogen to produce a protective immune response against *Staphylococcus aureus* infection using an animal (mouse or rat) model. The hybrid polypeptides having therapeutic and diagnostic applications, such as being used to provide protective immunity against *S. aureus* infection, being used to generate antibodies to detect the presence of *S. aureus*, and being used to generate therapeutic antibodies that target *S. aureus*.

Sequence 645 AA;

Query Match	12.6%	Score 94;	DB 9;	Length 645;
Best Local Similarity	25.0%	Pred. No. 2.2;		
Matches 40;	Conservative	28;	Mismatches 58;	Indels 34;
				Gaps 7;

Qy		2 RVVTI QNGKEMSSIV-----SREDFILPVYKGLEKGYQFDGWBSIFGEGKDAGY 54     :         :         :     :
Db		401 RVRTVSKDAKNRSRLIPPYIEGTKTYDAIVKVHVHTID-----YEQY--H H 445

	QY	V I N L S K O T F I P V P F K C I E E K B E E N K P T F D V - - - - S K G C O N P Q V H S Q L N E S H R K E D L Q	109
	Db	V I I I D I D O A F T A N T D K S M K G E Q Q O N S A K E A T P A T P S T F S V E K E S Q K Q S Q K D O D N K Q	505
	QY	- - - - R E E H S Q K S D S T K V T - A T V L D K N N I S S K S T T N P N K	144
	Db	L P S V E K E N D A S S E S G K D T P A T K P T K G E V B S S T T - - P T K	543

Search completed: April 24, 2006, 14:50:26  
Job time : 92.3624 secs

**THIS PAGE BLANK (USPTO)**  
**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 93.5369 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGEVSELKPHRVTVTQNG.....ATVLDKXNISKSTTNNPK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	799	100.0	Q9AHT5_STRPN	Q9Aht5 streptococc
2	799	100.0	Q97RY6_STRPN	Q97ry6 streptococc
3	796	99.6	Q8DQP7_STRPN	Q8dqp7 streptococc
4	795	99.5	Q9S4M8_STRPN	Q9s4m8 streptococc
5	129.5	16.2	Q4XUI6_PLACH	Q4xui6 plasmodium
6	118	14.8	Q8CPK8_STAEP	Q8cpk8 staphylococ
7	112.5	14.1	Q9SP15_PLAFA	Q9sp15 plasmodium
8	112.5	14.1	Q9U6C4_PLAFA	Q9u6c4 plasmodium
9	111.5	14.0	Q25706_PLAFA	Q25706 plasmodium
10	110.5	13.8	Q25705_PLAFA	Q25705 plasmodium
11	110.5	13.8	Q5HQ11_STAEP	Q5hq11 staphylococ
12	110	13.8	Q9U0G0_PLARE	Q9u0g0 plasmodium
13	109.5	13.7	Q5OVJ0_ENTHI	Q5ovj0 entamoeba h
14	109	13.6	Q26019_PLAFA	Q26019 plasmodium
15	109	13.6	Q71436_PLAF7	Q71436 plasmodium
16	108.5	13.6	Q77355_PLAF7	Q77355 plasmodium
17	108.5	13.6	Q90784_CHICK	Q90784 gallus gall
18	108.5	13.6	Q512T7_ENTHI	Q512t7 entamoeba h
19	108	13.5	Q25995_PLAFA	Q25995 plasmodium
20	108	13.5	Q81J55_PLAF7	Q81j55 plasmodium
21	107.5	13.5	Q9FJK9_ARATH	Q9fjk9 arabidopsis
22	107	13.4	Q815F3_PLAF7	Q815f3 plasmodium
23	106.5	13.3	Q501X8_ENTHI	Q501x8 entamoeba h
24	106.5	13.3	Q5V9M0_PLAXN	Q5v9m0 plasmodium
25	106.5	13.3	Q5W5T1_TETPY	Q5w5t1 tetrahymena
26	106	13.3	Q6BRW2_DBBHA	Q6brw2 debaromyce
27	105	13.1	Q4YMU4_PLABE	Q4ymu4 plasmodium
28	105	13.1	Q96229_PLAF7	Q96229 plasmodium
29	105	13.1	Q54GSI_DICDI	Q54gsi dictyosteli
30	104.5	13.1	Q4XYV2_PLABE	Q4xyv2 plasmodium
31	103.5	13.0	Q44016_DICDI	Q44016 dictyosteli

32	103	12.9	467	2	Q59PE2_CANAL	Q59pe2 candida alb
33	103	12.9	467	2	Q59PL2_CANAL	Q59pl2 candida alb
34	102.5	12.8	674	2	Q7RLI7_PLAYO	Q7rli7 plasmodium
35	102.5	12.8	2081	2	Q9LH98_ARATH	Q9lh98 arabidopsis
36	102	12.8	540	2	Q54MT2_DICDI	Q54mt2 dictyosteli
37	102	12.8	954	2	Q6HNR0_BACHK	Q6hnr0 bacillus th
38	102	12.8	1603	2	Q5SA14_DICDI	Q5sa14 dictyosteli
39	102	12.8	1603	2	Q8T216_DICDI	Q8t216 dictyosteli
40	101.5	12.7	382	2	Q9V7J0_DROME	Q9v7j0 drosophila
41	101.5	12.7	500	2	Q6BGL7_PASTE	Q6bgl7 paramescium
42	101.5	12.7	556	2	Q9V7I9_DROME	Q9v7i9 drosophila
43	101.5	12.7	785	2	Q9GQ82_DROME	Q9gq82 drosophila
44	101.5	12.7	1042	2	Q7REQ7_PLAYO	Q7req7 plasmodium
45	101	12.6	329	2	Q9NFV9_PLAFA	Q9nfv9 plasmodium

#### ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN  
ID Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5;  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
EX MEDLINE=21116976; PubMed=11179332;  
RX DOI=10.1128/IAI.69.3.1593-1598.2001;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.B., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Maguire H.R., Tuomanen E.,  
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
RA Langermann S., Johnson S., Koenig S.;  
RT "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 2S8T.  
DR MEROPS; S08.064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPTXG\_anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Protease.

FT NON TER 1 1  
SQ SEQUENCE 2119 AA; 238227 MW; 517F9B7F6B960A6A CRC64;  
Query Match 100.0%; Score 799; DB 2; Length 2119;  
Best Local Similarity 100.0%; Pred. No. 2.5e-48;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 1932 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKK 1991  
QY 61 DAGYVINLSKDTFIKPVFKKIEBKKEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQR 120  
DB 1992 DAGYVINLSKDTFIKPVFKKIEBKKEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQR 2051  
QY 121 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 154  
DB 2052 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 2085

## RESULT 2

Q9TRY6\_STRPN  
ID Q9TRY6\_STRPN PRELIMINARY; PRT; 2140 AA.  
AC Q9TRY6;  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-WAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Serine protease, subtilase family.  
GN OrderedLocusNames=SP0641;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAINS-ATCC BAA-334 / TIGR4;  
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,  
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";  
RL Science 293:498-506(2001).  
DR EMBL; AE007373; AAK74791.1; -; Genomic\_DNA.  
DR PIR; P95074; P95074.  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08.064; -.  
DR TIGR; SP0641; -.  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0004289; F:peptidase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.

DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS0847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
KW Cell wall; Complete proteome; Protease.  
SQ SEQUENCE 2140 AA; 240426 MW; F7A4AD8E2938B334 CRC64;  
Query Match 100.0%; Score 799; DB 2; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.6e-48;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 1953 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKK 2012  
QY 61 DAGYVINLSKDTFIKPVFKKIEBKKEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQR 120  
DB 2013 DAGYVINLSKDTFIKPVFKKIEBKKEENKPTFDVSKKDNPPQVNHSQLNESHKEDLQR 2072  
QY 121 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 154  
DB 2073 EHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 2106  
RESULT 3  
Q8DQF7\_STRR6  
ID Q8DQF7\_STRR6 PRELIMINARY; PRT; 2144 AA.  
AC Q8DQF7;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Cell wall-associated serine proteinase PrtA (SC 3.4.21.-).  
GN Name=prtA; OrderedLocusNames=sp0551;  
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=171101;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21429245; PubMed=11544234;  
RX DOI=10.1128/JB.183.19.5709-5717.2001;  
RA Hoskins J., Albom W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,  
RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,  
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,  
RA Glass J.I.;  
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";  
RL J. Bacteriol. 183:5709-5717(2001).  
DR EMBL; AE008434; AAK9365.1; -; Genomic\_DNA.  
DR PIR; A97942; A97942.  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08.064; -.  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.



DR Pfam; PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase S8; 1.  
 DR Pfam; PF05922; Subtilisin N; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
 DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Cell wall; Signal.  
 FT SIGNAL 1 19 Potential.  
 FT CHAIN 20 2144 Cell wall-associated serine proteinase  
 FT PrtA.  
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.6%; Score 796; DB 2; Length 2144;  
 Best Local Similarity 99.4%; Pred. No. 4.2e-48;  
 Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGLEKGYQFDGWEISGFEKK 60  
 |||||  
 Db 1957 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGLEKGYQFDGWEISGFEKK 2016  
 |||||

Qy 61 DAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 120  
 |||||  
 Db 2017 DAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 2076  
 |||||

Qy 121 EHSQKSDSTKQVTVATVLDKNNISSKSTNNPNK 154  
 |||||  
 Db 2077 EDHSQKSDSTKQVTVATVLDKNNISSKSTNNPNK 2110  
 |||||

RESULT 4  
 ID Q9S4M8\_STRPN PRELIMINARY; PRT; 2144 AA.  
 AC Q9S4M8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Cell wall-associated serine proteinase precursor PrtA.  
 GN Name=prtA;  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3.B;  
 RX MEDLINE=21585565; PubMed=11728722;  
 RA Beche G., Nau R., Wellner A., Hakenbeck R., Reinert R.R., Heinz H.P.,  
 RA Zysk G.;  
 RA "The cell wall-associated serine proteinase PrtA: a highly conserved  
 RT virulence factor of Streptococcus pneumoniae.";  
 RL FEMS Microbiol. Lett. 205:99-104(2001).  
 DR EMBL; AF127143; AAD48399.1; -; Genomic\_DNA.  
 DR HSSP; P00782; 2SER.  
 DR MEROPS; S08.064; -;  
 DR GO; GO:0009986; C:cell surface; IEA.  
 DR GO; GO:0005618; C:cell wall; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0042802; F:protein self binding; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR010435; DUP1034.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR003137; PA.  
 DR InterPro; IPR000209; Pept\_S8\_S53.  
 DR InterPro; IPR010259; Prot\_inf\_S8A.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF06280; DUF1034; 1.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF02225; PA; 1.  
 DR Pfam; PF00082; Peptidase S8; 1.  
 DR Pfam; PF05922; Subtilisin N; 1.

DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
 DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
 KW Cell wall; Signal.  
 FT SIGNAL 1 19 Potential.  
 FT CHAIN 20 2144 Cell wall-associated serine proteinase  
 FT PrtA.  
 SQ SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;

Query Match 99.5%; Score 795; DB 2; Length 2144;  
 Best Local Similarity 98.7%; Pred. No. 5e-48;  
 Matches 152; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGLEKGYQFDGWEISGFEKK 60  
 |||||  
 Db 1957 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGLEKGYQFDGWEISGFEKK 2016  
 |||||

Qy 61 DAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 120  
 |||||  
 Db 2017 DAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 2076  
 |||||

Qy 121 EHSQKSDSTKQVTVATVLDKNNISSKSTNNPNK 154  
 |||||  
 Db 2077 EDHSQKSDSTKQVTVATVLDKNNISSKSTNNPNK 2110  
 |||||

RESULT 5  
 ID Q4XUI6\_PLACH PRELIMINARY; PRT; 300 AA.  
 AC Q4XUI6;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN ORFNames=PC000286.03.0;  
 OS Plasmodium chabaudi.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5825;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,  
 RA Beriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,  
 RA James K., Rutherford K., Harris B., Harris D., Churcher C.,  
 RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,  
 RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,  
 RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;  
 RT "A comprehensive survey of the Plasmodium life cycle by genomic,  
 RL transcriptomic, and proteomic analyses.";  
 RL Science 307:82-86(2005).  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; CAATJ01003049; CAH79425.1; -; Genomic\_DNA.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 4.  
 DR PRINTS; PR00320; GPROTEINRPT.  
 DR ProDom; PD0000018; WD40; 3.  
 DR SMART; SM00320; WD40; 5.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 DR PROSITE; PS00082; WD\_REPEATS\_2; 4.  
 DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 KW Hypothetical protein; Repeat; WD repeat.  
 FT NON TER 1  
 SQ SEQUENCE 300 AA; 34469 MW; 8EDB4E512AFB1945 CRC64;

Query Match 16.2%; Score 129.5; DB 2; Length 300;  
 Best Local Similarity 26.8%; Pred. No. 0.21;  
 Matches 40; Conservative 31; Mismatches 45; Indels 33; Gaps 7;

Qy 9 KPHRVTVTIQNGKMSSTIVSE-----EDFILPVYKGLEKGYQFDGWEISGFEKK 60

```

Db 130 KVIHTQSPNGKFTASSPDKSIRWSGIDGTYLVAVRHGVGPAYKI-AWSI----- 181
QY 61 DAGYVINLSKDTFIK-----PVPKIEEKE---FENKPTFVDSKKNDPQVNHSQL 109
Db 182 DNNYIVSCSDSTLKLWRINHLVPLLLKKEENAEQTKDEK-----NQKENPQ-NNDP 235
QY 110 NESHRKEDLQREHSQKSDSTKDVATVTL 138
Db 236 NDEANSEKKKNEKEKNDTKTKTKTL 264

RESULT 6
Q8CPK8 STAEPP
ID Q8CPK8 STAEPP PRELIMINARY; PRT; 775 AA.
AC Q8CPK8
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Penicillin-binding protein 1.
GN OrderedLocusNames=SE0856;
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 12228;
RX PubMed:12950922; DOI=10.1046/j.1365-2958.2003.03671.x;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao S.-G., Wang H.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593 (2003).
DR EMBL; AB016746; AA04453.1; -; Genomic_DNA.
DR HSSP; P14677; 1QME.
DR GO; GO:0008658; P:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR005311; PBP_dimer.
DR InterPro; IPR012338; PBP_tpept_fold.
DR InterPro; IPR001460; Pencil_bind_tpept.
DR Pfam; PF03793; PASTA; 2.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR SMART; SM00740; PASTA; 2.
KW Complete proteome.
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;

Query Match 14.8%; Score 118; DB 2; Length 775;
Best Local Similarity 27.2%; Pred. No. 3.9;
Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;

QY 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQ----- 51
Db 615 DSVNAQSLKP-----ITIGNGKIQKQSVKSGTKVLPHSKVWMLTGDGLTMP-DWTGWTKE 669
QY 52 EISGFE-----GKQDAGYVIN--LSKDTFTKPVFKIEEKEENKPTFDVS-----KK 98
Db 670 DVLAFEDLTIKIVSTKNGSFTVNTQISGKQIIK-----NKDKIEVLSAEDT 716
QY 99 KNPQVNHSQLNESHRKEDLQREHSQKSDSKDTVATVLDKNNTSKSTNN 151
Db 717 DDQKEKTEDSDSNKSKKDKADEHSNTSSSTNN-----DKSNADSKNDSDD 763

RESULT 7
Q95P15 PLAF6
ID Q95P15 PLAF6 PRELIMINARY; PRT; 361 AA.
AC Q95P15
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

Query Match 14.8%; Score 118; DB 2; Length 379;
Best Local Similarity 23.6%; Pred. No. 4.5;
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;

QY 9 KPHRVTTVIQNGKMSSTIVSEDF-----ILPVYKGELEKGYQFD-GWEISGFP- 56
Db 171 KPSRLNLFSRKTKYASQV--EKDYERAKNAYKANQAVLKAKESASYDYILGWFFGGV 228

```

```

DE Merozoite surface protein 3 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVO;
RX MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;
RA Hsaeda H., Saul A., Reece J.J., Kennedy M.C., Long C.A., Miller L.H.,
RA Stowers A.W.;
RT "Merozoite surface protein 3 and protection against malaria in Aotus
RT nancymai monkeys.";
RL J. Infect. Dis. 185:657-664 (2002).
DR EMBL; AY044180; AAK94780.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAN.
DR Pfam; PF07133; Merozoite_SPAN; 1.
KW Merozoite.
FT CHAIN <1> >361 merozoite surface protein 3.
FT NON_TER 1 1
FT NON_TER 361 361
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 14.1%; Score 112.5; DB 2; Length 361;
Best Local Similarity 22.4%; Pred. No. 4.3;
Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;

QY 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQ----- 47
Db 147 ETGE-----RNSGRNPNFYTKTKB-----YAGKVKDYERAKNAYKANQAV 187
QY 48 -----FD---GWEISGFP---EGKIDAG-----YVINLSKDTFTKPVFKIEEKEEN 89
Db 188 LKAEASSYDYLGWFFGGVPEHKGKGNMLSHLYSVSKKKNISKENDVDVLDE-KEEEA 246
QY 90 KPTFDVSKKNDPQVNHSQLNESHRKEDLQREHSQKSDSKDTVATVLDKQ----- 141
Db 247 EETEEELKEKNEETESEISEDEHEEHEEKEEENDKKQEKQESQNNENNDQKKDME 306
QY 142 --NISSKSTNN 151
Db 307 AQNLISKNNNN 318

RESULT 8
Q9U6C4 PLAF6
ID Q9U6C4 PLAF6 PRELIMINARY; PRT; 379 AA.
AC Q9U6C4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polymorphic antigen.
GN Name=MSP-3;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PCCL/HN;
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;
RL Submitted (SFP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188190; AAF04099.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAN.
DR Pfam; PF07133; Merozoite_SPAN; 1.
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1FP9D5F25 CRC64;

Query Match 14.1%; Score 112.5; DB 2; Length 379;
Best Local Similarity 23.6%; Pred. No. 4.5;
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;

QY 9 KPHRVTTVIQNGKMSSTIVSEDF-----ILPVYKGELEKGYQFD-GWEISGFP- 56
Db 171 KPSRLNLFSRKTKYASQV--EKDYERAKNAYKANQAVLKAKESASYDYILGWFFGGV 228

```

QY 57 -EGKKDAG-----YVNLKSDTFIKPVFKIEEKEEENKPTFDVSKKKNQPNVHNSQLN 110  
 Db 229 PEHKKENMLSHLYVSSKDKENISKENDVDLDE-KEEBAETEEBELEKEEETESIS 287  
 QY 111 -----ESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNN 151  
 Db 288 EDEEBEKEEENKKEKEKEQSNENNDDQKDMAE-----QNLISKNNNN 336

## RESULT 9

Q25706 PLAFPA  
 ID Q25706\_PLAFPA PRELIMINARY; PRT; 379 AA.  
 AC Q25706  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Polymorphic antigen.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]

## NUCLEOTIDE SEQUENCE.

RP MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;  
 RA McColl D.J., Anders R.F.;  
 RT "Conservation of structural motifs and antigenic diversity in the  
 RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
 RL Mol. Biochem. Parasitol. 90:21-31(1997).  
 DR EMBL; U08852; AAC47832.1; -; Unassigned\_DNA.  
 DR InterPro; IPR010784; Merozoite SPAM.  
 DR Pfam; PF07133; Merozoite SPAM; 1.  
 DR SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;

Query Match 14.0%; Score 111.5; DB 2; Length 379;  
 Best Local Similarity 23.6%; Pred. No. 5.3;  
 Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;

QY 9 KPHRVTVTQNGKEMSSITVSEEDP-----ILPVYKGELEKGYQPD-GWEISGF- 56  
 Db 171 KPSRNLFSRKTYYAEQV--EKDYERAKNAYQKANQAVLKAKESASYDYILGWFGGV 228

QY 57 -EGKKDAG-----YVNLKSDTFIKPVFKIEEKEEENKPTFDVSKKKNQPNVHNSQLN 110  
 Db 229 PEHKKENMLSHLYVSSKDKENISKENDVDLDE-KEEBAETEEBELEKEEETESIS 287

QY 111 -----ESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNN 151  
 Db 288 EDEEBEKEEENKKEKEKEQSNENNDDQKDMAE-----QNLISKNNNN 336

## RESULT 10

Q25705 PLAFPA  
 ID Q25705\_PLAFPA PRELIMINARY; PRT; 379 AA.  
 AC Q25705  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Polymorphic antigen.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]

## NUCLEOTIDE SEQUENCE.

RP MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;  
 RA McColl D.J., Anders R.F.;  
 RT "Conservation of structural motifs and antigenic diversity in the  
 RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
 RL Mol. Biochem. Parasitol. 90:21-31(1997).  
 DR EMBL; U08851; AAC47831.1; -; Unassigned\_DNA.  
 DR InterPro; IPR010784; Merozoite SPAM.  
 DR Pfam; PF07133; Merozoite SPAM; 1.  
 DR SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 13.8%; Score 110.5; DB 2; Length 379;  
 Best Local Similarity 25.0%; Pred. No. 14;  
 Matches 42; Conservative 27; Mismatches 66; Indels 33; Gaps 8;

Best Local Similarity 22.1%; Pred. No. 6.3;  
 Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9;  
 QY 1 DTGVSSELKPHRVTVTQNGKEMSSITVSEEDPILPVYKGELEKGYQ----- 47  
 Db 167 ETGB-----RNSNNPNTYTKE-----YAGKVKDYERAKNAYQKANQAV 207  
 QY 48 -----FD---GWISGF--EGKKDAG-----YVNLKSDTFIKPVFKIEEKEEEN 89  
 Db 208 LKAKEASSYDYLGWFGGVPEHKENMLSHLYVSSKDKENISKENDVDLDE-KBEEA 266  
 QY 90 KPTFDVSKKKNQPNVHNSQLN-----ESHKEDLQREHHSQKSDSTKDVAT 136  
 Db 267 ETEEBELEKEEETESISEDEEBEKEEENKKEKEKEQSNENNDDQKDMAE- 325  
 QY 137 VLDKNNISSKSTNN 151  
 Db 326 ----QNLISKNNNN 336

## RESULT 11

Q5HQ11 STABO  
 ID Q5HQ11 STABO PRELIMINARY; PRT; 775 AA.  
 AC Q5HQ11  
 DT 10-MAY-2005 (TRENBLrel. 30, Created)  
 DT 10-MAY-2005 (TRENBLrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TRENBLrel. 30, Last annotation update)  
 DE Penicillin-binding protein 1.  
 GN Name=pbpl; OrderedLocName=SERP0746;  
 OS Staphylococcus epidermidis (strain ATCC 35984 / RP62A).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=176279;  
 RN [1]

## NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RP PubMed=1574886; DOI=10.1128/JB.187.7.2426-2438.2005;  
 RA Gill S.R., Fouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T.,  
 RA Ravel J., Paulsen I.F., Kolonay J.F., Brinkac L.M., Seaman M.J.,  
 RA Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S.,  
 RA Haft D.H., Vamathevan J.J., Khouri H., Utterback T.R., Lee C.,  
 RA Dmitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H.,  
 RA Hance I.R., Nelson K.E., Fraser C.M.;  
 RT "Insights on evolution of virulence and resistance from the complete  
 RT genome analysis of an early methicillin-resistant Staphylococcus  
 RT aureus strain and a biofilm-producing methicillin-resistant  
 RT Staphylococcus epidermidis strain.";  
 RL J. Bacteriol. 187:2426-2438(2005).  
 DR EMBL; CP000029; AAWS4126.1; -; Genomic\_DNA.  
 DR TIGR; SERP0746; -

DR GO; GO:0008658; P:penicillin binding; IEA.  
 DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.  
 DR InterPro; IPR005543; PASTA.  
 DR InterPro; IPR005311; PBP dimer.  
 DR InterPro; IPR001460; PenC\_bind\_tpept.  
 DR Pfam; PF03793; PASTA; 2.  
 DR Pfam; PF03717; PBP\_dimer; 1.  
 DR Pfam; PF00905; Transpeptidase; 1.  
 DR SMART; SM00740; PASTA; 2.  
 DR KW Complete proteome.  
 DR SQ SEQUENCE 775 AA; 86352 MW; B9395893E0043694 CRC64;

Query Match 13.8%; Score 110.5; DB 2; Length 775;  
 Best Local Similarity 25.0%; Pred. No. 14;  
 Matches 42; Conservative 27; Mismatches 66; Indels 33; Gaps 8;

QY 1 DTGVSSELKPHRVTVTQNGKEMSSITVSEEDPILPVYK-----GLEKGYQPDG--- 51  
 Db 615 DSVNAQSLKP-----ITIGNKGKIQKQSVSGTKVPLPHSKVLMVMTDGLTWP-DMTGWTK 669  
 QY 52 EISQFB-----GKKDAGVYIN--LSKDTFIKPVFK-----KIEEKEEENKPTFD 94  
 Db 670 DVLAPEDLTKIKVSTKNGFVTVNQSIKGIKKDKIEVSLSAEDTDDOEKTDDESSD 729  
 QY 95 VSKKKNQPNVHNSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNN 142



```
Db 225 FGGVPEHKKENMLSHLYSSKDKENISKENDVDLDE-KREABETEERLEEKNEEET 283
QY 106 HSQLN-----ESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNN 151
Db 284 ESEISEDEEBEERKEENKKEQKESQENNDQKDEA-----QNLISKQNNN 337

RESULT 15
Q81436 PLAP7
ID Q81436 PLAP7 PRELIMINARY; PRT; 3008 AA.
AC Q81436;
DT 01-WAR-2003 (TREMBlrel. 23, Created)
DT 01-WAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-WAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein PFE0325w.
GN Name=PFE0325w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=3D7;
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=3D7;
RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL929351; CAD51431.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 3008 AA; 356025 MW; 60BCBBEB15C599B4 CRC64;

Query Match 13.6%; Score 109; DB 2; Length 3008;
Best Local Similarity 32.4%; Pred. No. 74;
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 64 YVINLSK-----DTFKPVFKKIEKKKEENKPTFDVSKKQDPQVNHSHQ----LNESHKRE 116
Db 2310 YDIELSKIEKFGKSGIPVFTD-EENKKEENKN-EVNNKKEENKKEENKNEVNNKKE 2366

QY 117 DLQREH-----SQKSDSTKDVATVLDKNNISSK-----STNNPNK 154
Db 2367 ENKKEENKKEENKKEENKKEENKNEVNNKKEENKKEENKKEENK 2414

RESULT 16
O77355 PLAP7
ID O77355 PLAP7 PRELIMINARY; PRT; 600 AA.
AC O77355;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-2000 (TREMBlrel. 15, Last sequence update)
DT 01-WAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein MAL3P4.20.
GN Name=MAL3P4.20; Synonym=PPC0465c;
```

```
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sultston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum.";
RL Nature 400:532-538(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sultston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
DR EMBL; AL008970; CAAL15610.2; -; Genomic_DNA.
DR PIR; T18467; T18467.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0006397; P:mRNA processing; IEA.
DR InterPro; IPR002483; PWI.
DR Pfam; PF01480; PWI; 1.
DR SMART; SM00311; PWI; 1.
KW Hypothetical protein; Lyase.
SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;

Query Match 13.6%; Score 108.5; DB 2; Length 600;
Best Local Similarity 29.3%; Pred. No. 14;
Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;

QY 27 IVSEEDTLPVY-----KGELEKGVQPDGWEISGFEGKK-----DAGVYVNLSKDTFKPV 77
Db 60 ILGFEDDILYCYICSQLKQSEKK---DGEEDKVLNKKLKLNTLFGKNGKSDIFIEL 116

QY 78 FKXI--BEKKEE-----ENKPTFDVSK-KQDPQVNHSQLNE-----SHRK 115
Db 117 LELLINBEKKEHIADTLNENK-TNDIKKVENENINENYVNMKDISNKKDHVSHQN 175

QY 116 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTNN 150
Db 176 EHNINNNVNLKKEBYTDIQDRKHGRSLSKQSDSYKKRPFNKRTKSTIER-SLSNRYDE 234

QY 151 NPNK 154
Db 235 KTNK 238

RESULT 17
Q90784 CHICK
ID Q90784_CHICK PRELIMINARY; PRT; 1038 AA.
AC Q90784;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
```

```
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Claustrin.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]_TaxID=9031;
RP NUCLEOTIDE SEQUENCE.
RX TISSUE=Brain;
RA MEDLINE=94157526; PubMed=7906711;
RA Burg M.A., Cole G.J.;
RT "Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is
RT structurally related to MAP1B.";
RL J. Neurobiol. 25:1-22(1994).
DR EMBL; X67778; CAA47988.1; -; mRNA.
DR PIR; JC5497; JC5497.
DR Ensemble; ENSGALG0000014999; Gallus gallus.
SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;

Query Match 13.6%; Score 108.5; DB 2; Length 1038;
Best Local Similarity 28.6%; Pred. No. 26;
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

QY 21 KEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGEGKQAGYVNLKDTPIKPVK- 79
Db 595 KPEETKIIVAKRDV-----TTKEQLGKSTSEKQASEKQDVKPKVSKVKVEKVA 646

QY 80 KLEEKKEENKPTFDVSKKKNQPNVHNSQLNESHKEDLQRE-----BHSQKSDSTKDV 133
Db 647 KPEKKKDEKPKVEKVEKPLI---KKEKPKKEIKVEKVEKVEKVEKVEKVEK 702

RESULT 18
Q512T7 ENTH1
ID Q512T7_ENTH1 PRELIMINARY; PRT; 1069 AA.
AC Q512T7_
DT 13-SEP-2005 (Tremblrel. 31, Created)
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
DE Fimbriae-associated protein, putative.
GN ORFNames=89.t00007;
OS Entamoeba histolytica HW-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=HM-1:IMSS;
RA Loftus B., Anderson I., Davies R., Alenmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoest U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harria D.,
RA Jagals K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.B., Bhattacharya S., Lohia A.,
RA Foster P.G., Sichert-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFB01000328; EAL47849.1; -; Genomic DNA.
SQ SEQUENCE 1069 AA; 120249 MW; 1D38E4F0A8759CB4 CRC64;

Query Match 13.6%; Score 108.5; DB 2; Length 1069;
Best Local Similarity 25.5%; Pred. No. 27;
Matches 42; Conservative 34; Mismatches 60; Indels 29; Gaps 7;
```

```
QY 4 EVSELKPHRVTVTIQNG-KEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGEGKDA 62
Db 138 EVSKDNVNESSTLTNGEKKLSLNCNEQD-----ELQSKSSSTD--NKNDKDE 187

QY 63 GYVNLKSDTFIKPVFKKIEKKEENK-----PTFDVSKKKNQ-----VNHSQLN 110
Db 188 IHFDVFLPKNEKEISMEIESSKTEEKSNIQLPSLNLSEKGNVESVEIAKVLKKSNS 247

QY 111 ESHRKEDLQREHS-QKSDSTKD-----VTATVLDKKNISSKSTT 149
Db 248 NNSGEEDKQDEEVSCFKPDSQEEKKEMIKAEVSNKQKVKDKSTT 292

RESULT 19
Q25995 PLAF7
ID Q25995_PLAF7 PRELIMINARY; PRT; 354 AA.
AC Q25995;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=NF54;
RA MEDLINE=95198774; PubMed=7891748; DOI=10.1016/0166-6851(94)00149-9;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.P.;
RT "Molecular variation in a novel polymorphic antigen associated with
RT plasmodium falciparum merozoites.";
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]_
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=NF54;
RA MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L28825; AAC09377.1; -; Genomic DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 13.5%; Score 108; DB 2; Length 354;
Best Local Similarity 23.2%; Pred. No. 8.8;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 17 IQNGKEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFP--EGKQAG-----YVINLS 69
Db 181 VLKAKEASS-----YDYIL-----GWFEFGGVPEHKKEENMLSHLYVSSKD 221

QY 70 KDTFKPKVPFKIEKKEE-----ENKPTFDVSKKKNQPNVHNSQLNESHKRE 116
Db 222 KENISKENDVDLDKEEAEETEEELKEKNEETSEISEDEEEEEEEKEEENDKKK 281

QY 117 DLQREHSQKSDSTKDVATVATVLDKKNISSKSTTNN 151
Db 282 EQEKEQSNNNDQKDMA-----QLISKQNNN 311

RESULT 20
Q81J55 PLAF7
ID Q81J55_PLAF7 PRELIMINARY; PRT; 354 AA.
AC Q81J55;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Merozoite surface protein 3.
GN ORFNames=PF10_0345;
OS Plasmodium falciparum (isolate 3D7).
```







```

RESULT 27
Q4YMU4_PLABE
ID Q4YMU4_PLABE PRELIMINARY; PRT; 662 AA.
AC Q4YMU4;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE RNA binding protein, putative.
GN ORFNames=PB001104.03.0;
OS Plasmodium berghei.
OC Eukaryote; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
CC NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolj T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides
RA James K., Rutherford K., Harris D., Harris D., Churcher C.,

```

RESULT	28			
O96229	PLAF7			
ID	O56229_PLAF7	PRELIMINARY;	PRT;	951 AA.
AC	O96229;			
DT	01-MAY-1999	(TREMBLrel. 10, Created)		
DT	01-WAR-2003	(TREMBLrel. 23, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE	Hypotheothical protein PFB0680w.			
GN	Name=PFB0680w;			
OS	Plasmodium falciparum (isolate 3D7).			

```
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AF001410; RAC71925.2; -; Genomic_DNA.
KW PIR; B71609; B71609.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 112486 MW; AC8DB89358A84F4F CRC64;

Query Match 13.1%; Score 105; DB 2; Length 951;
Best Local Similarity 24.4%; Pred. No. 42;
Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;

Qy 1 DTGVSSELKPHRYT-VTTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGFE 57
Db 129 EKKINKSLHRLNLSQSGK-----NEQDI-----NKVEKGRQ-----DISNSNAE 171
Qy 58 GKGDAGYVNLKSDTFIKPVFKKIEKKE-----EENKPTFD-----VSKKDNPPQ 103
Db 172 NKDD-----VKEGVKELEKKEEIKSDHKVKNKSDDKHKVKNKSDDKH 219
Qy 104 VNHSQNLNESHKEDLQR-EHHSQKSDSTKDVATVTLVDKNNISSTKTTNPNK 154
Db 220 VEENKKSDDHKTEEVKVEEHEDEEE-----DKKEKSENKNDENK 262

RESULT 29
Q54GS1 D1C1D1 PRELIMINARY; PRT; 1550 AA.
AC Q54GS1;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=DD0188660;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Degany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangaveilu M., Johnson D., Knights A.,
RA Loulsegged H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J.J., Winckler T., Tanaka Y.,
RA Shaulyuk G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzner M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFI01000221; EAL62484.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1550 AA; 174527 MW; D6D5DFP547DC48B6 CRC64;

Query Match 13.1%; Score 105; DB 2; Length 1550;
Best Local Similarity 25.4%; Pred. No. 70;
Matches 46; Conservative 24; Mismatches 59; Indels 52; Gaps 6;
```

```
Qy 10 PHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEK--GYQFDGWEISGFGKDGAGYVI- 66
Db 1328 PHK-----LKKGNRCSSSSSLNDSILPTLEDHLAKTYSPKVDMENTSFDDTLDLKELIS 1383
Qy 67 -----NLKSDTFIKPVFKKIEKKEEENKPTFDVS----- 96
Db 1384 TDELNNGNNNNNEDNNENPFSNIKENSVKVGQVPVQOEIEKEKEKENIIGEDVEKETS 1443
Qy 97 -----KKONPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVTLVDKNNISKSTTN 150
Db 1444 IEQKEIKKDDNNNNNIVEKEEEIKDNIKBEIKQDDSNNNKBEI-----NN--DNSTTN 1494
Qy 151 N 151
Db 1495 N 1495

RESULT 30
Q4YVY2 PLABE PRELIMINARY; PRT; 736 AA.
AC Q4YVY2;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=PB000556.02.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAT1002190; CAH97824.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 736 AA; 85816 MW; A1315CF2D97A6905 CRC64;

Query Match 13.1%; Score 104.5; DB 2; Length 736;
Best Local Similarity 23.0%; Pred. No. 34;
Matches 41; Conservative 34; Mismatches 60; Indels 43; Gaps 8;

Qy 17 IONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVNLKSDT----- 72
Db 401 INNNDSSS-----GKKNAIPIFLKCKKNNKCDTIETNPWINKAGDLVKKKESLTL 457
Qy 73 FIKPVFKKIEKKEE-----ENKPTFDVSKKDNPPVNH-----SOLNESHK----- 114
Db 458 FSSPLHKKIEKKEKENVNDYENAKVENIKKNNKGRIKHPLKIEFYFNEENKINTPKPPDD 517
Qy 115 ----KEDLQREHSHQKSD-----STKDVATV--VLDK-----NNISKSTTNPNK 154
Db 518 IISANENPKTNSPFSKSDILDTSVNNNANAYEIFEKGDKNKNSKVDSNIIDNDK 575

RESULT 31
Q44016 D1C1D1 PRELIMINARY; PRT; 325 AA.
AC Q44016;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE G5 ORF.
OS Dictyostelium discoideum (slime mold).
```



```
Db 251 TEEVPIIE-----NFONYD--NDEDKKE--VVKSKNE--NQNTTKMESKPKPKPK 299
Qy 92 TFDVSKKKNPOVN-----HSQLNESHKEDLQREHS 124
Db 300 SEASATKVKPQPRKQKQKPLSBETVDLTDLDLDDDFKDLQLELLEBEEQPKQOI 359
Qy 125 QKSDSTKDV-----TATVLDKNNISKSTNNPNK 154
Db 360 KETKSNQSIQKPSPTIEVDPIAFNDSDEDPEDHFTGIKIDEGNNSNSSNNNK 417

RESULT 34
Q7RLE7 PLAYO
ID Q7RLE7_PLAYO PRELIMINARY; PRT; 674 AA.
AC Q7RLE7_
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Notchless-related.
GN Name=PY02598;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP STRAIN=17XNL;
RC MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angluoli S.V., Suh B.B., Kooji T.W., Perte M.,
RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallon S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABL01000712; EAA22065.1; -; Genomic_DNA.
DR HSSP: P16649; 1ER3.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 4.
DR PROSITE: PS00678; WD_REPEATS_1; 2.
DR PROSITE: PS00082; WD_REPEATS_2; 6.
DR PROSITE: PS02094; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 674 AA; 76535 MW; E7521B469FE8E0F7 CRC64;

Query Match 12.8%; Score 102.5; DB 2; Length 674;
Best Local Similarity 23.6%; Pred. No. 43;
Matches 37; Conservative 27; Mismatches 58; Indels 35; Gaps 6;

Qy 9 KPHRVTVTQNGKMSSTIVS-----EDFILPVYKGELEKGYQFDGWEISGFEKK 60
Db 490 KVIHTQSPNGKFTASSFDKSIWMSGIDGTYLAVFRGHVPAYKI--AWSI----- 541
Qy 61 DAGYVINLSKDTPIK-----PVFKIE-----KKENKPTFDV--SKKDN 101
Db 542 DNNYIISQSDSTLKLWRINHLVPLLLKKEENGQPKSDPKNDQKNDQKDKQDKQDKDQ 601
Qy 102 PQVNSQLNESHKEDLQREHSQKSDSTKDVATVL 138
Db 602 KNDNNDHANNQEDGBEKKKKKKKKKKKKIKITLL 638
```

```
RESULT 35
Q9LH98 ARATH
ID Q9LH98_ARATH PRELIMINARY; PRT; 2081 AA.
AC Q9LH98_
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
RT and BAC clones."
RL DNA Res. 7:217-221(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002057; BAB03174.1; -; Genomic_DNA.
DR HSSP: P01096; 1HF9.
DR InterPro: IPR003605; DUF1216.
DR Pfam: PF06746; DUF1216; 2.
SQ SEQUENCE 2081 AA; 232852 MW; D3603E1F85EFPF29 CRC64;

Query Match 12.8%; Score 102.5; DB 2; Length 2081;
Best Local Similarity 27.4%; Pred. No. 1.5e+02;
Matches 43; Conservative 24; Mismatches 63; Indels 27; Gaps 6;

Qy 14 TTTTQNGKMSSTIVSBEEDFILPVYKGELEKGYQFDGWEISGFEKKD-----AG 63
Db 1659 TVEINGGBELSTERGSKD-----GKIEBGK--EGKENSTKEGSKDKKIEGMEGKEN 1708
Qy 64 YVINLSKDTPIKPVFKIEKKKEENKPF--TFDVSK-KDNQPNVNSQLNESHKEDLQ- 119
Db 1709 STKESKDGKINEIHGDKKATMEEGSKDGGTNGSKDQSKSKSVKVEINGKDDSLKDDSKN 1768
Qy 120 ---REHSQKSDSTKDVATVLDKNNISKSTNNPN 153
Db 1769 GDINEINNGKEDSVKDNVTIQTGNDNSLTNSSEPN 1805

RESULT 36
Q54MT2 DICDI
ID Q54MT2_DICDI PRELIMINARY; PRT; 540 AA.
AC Q54MT2_
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN OREName=DDB0186654;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugeng B., Kummerfeld S., Madera M., Konfortov B.A., Rivero P.,
RA Tungal B., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Bankier K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Faubrother P., Besany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
```





Search completed: April 24, 2006, 14:59:37  
Job time : 95.5369 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:50:52 ; Search time 16.1924 Seconds  
(without alignments)  
915.083 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773  
Perfect score: 799  
Sequence: 1 DTGEVSELKPHRVTTIQNG.....ATVLDKNNISSKSTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	799	100.0	2140	2	F95074	serine proteinase,
2	796	99.6	2144	2	A97942	metalloproteinase
3	110	13.8	558	2	T18467	hypothetical prote
4	108.5	13.6	1038	2	JC5497	claustrin - chicke
5	105	13.1	665	2	B71609	hypothetical prote
6	103.5	13.0	325	2	T18283	hypothetical prote
7	100	12.5	312	2	G81339	probable membrane
8	99	12.4	211	2	T25911	hypothetical prote
9	97.5	12.2	614	2	A84152	hypothetical prote
10	97.5	12.2	1345	2	S46817	hypothetical prote
11	97	12.1	385	2	T20410	hypothetical prote
12	97	12.1	988	2	T14188	hypothetical prote
13	96.5	12.1	540	2	D86432	hypothetical prote
14	96	12.0	535	2	T37189	hypothetical prote
15	96	12.0	2500	2	G71609	hypothetical prote
16	95.5	12.0	644	2	T47835	hypothetical prote
17	95	11.9	348	2	T37271	cyclicin II - human
18	95	11.9	622	2	A90570	lipoprotein (impor
19	95	11.9	1397	2	T10466	DNA topoisomerase
20	94.5	11.8	3724	2	T18427	hypothetical prote
21	94	11.8	210	2	T28771	hypothetical prote
22	93.5	11.7	456	2	T05612	hypothetical prote
23	93	11.6	219	2	B72291	hypothetical prote
24	91	11.4	253	2	T32879	hypothetical prote
25	90.5	11.3	629	2	G96542	hypothetical prote
26	90.5	11.3	670	2	T28391	ORF MSV230 hypothe
27	90.5	11.3	867	2	T27136	hypothetical prote
28	90.5	11.3	871	2	T27135	hypothetical prote
29	90.5	11.3	1332	2	S41552	probable transcrip

30	90.5	11.3	2401	2	T28676	rhoptry protein -
31	90	11.3	645	2	E89883	conserved hypothet
32	90	11.3	1202	1	S05362	probable DNA-direc
33	89.5	11.2	433	2	A89951	trigger factor (im
34	89.5	11.2	508	2	B11594	hypothetical prote
35	89.5	11.2	508	2	E86549	hypothetical prote
36	89.5	11.2	508	2	C72074	hypothetical prote
37	89.5	11.2	700	2	S67610	probable membrane
38	89.5	11.2	1016	2	T19006	ankyrin related pr
39	89	11.1	528	2	E96795	unknown protein P2
40	89	11.1	1888	2	T39009	hypothetical prote
41	88.5	11.1	301	2	T33068	hypothetical prote
42	88.5	11.1	371	2	A71683	hypothetical prote
43	88.5	11.1	443	2	S66040	serine-type D-Ala-
44	88.5	11.1	762	2	G88436	protein T04A8.13 [
45	88.5	11.1	791	2	T24435	hypothetical prote

## ALIGNMENTS

## RESULT 1

F95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: F95074  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heaton, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, R.; Hickey, E.K.; Holt, I.E.  
non, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95074  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2140 <KUR>  
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match	100.0%	Score	799;	DB	2;	Length	2140;
Best Local Similarity	100.0%	Pred. No.	1.3e-51;				
Matches	154;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	DTGEVSELKPHRVTTIQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGEGKK	60				
Db	1953	DTGEVSELKPHRVTTIQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGEGKK	2012				
Qy	61	DAGVYINLSKDTFKPVFKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQR	120				
Db	2013	DAGVYINLSKDTFKPVFKIEKKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQR	2072				
Qy	121	EEHSQKSDTKDVTATVLDKNNISSKSTNNPNK	154				
Db	2073	EEHSQKSDTKDVTATVLDKNNISSKSTNNPNK	2106				

## RESULT 2

A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; J. P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: A97942  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2144 <KUR>  
A;Cross-references: UNIPROT:Q8DQ7; UNIPARC:UPI00000E3490; GB:AE007317; PIDN:AAK99365.1;  
C;Genetics:  
A;Gene: pTnA  
C;Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 796; DB 2; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 2.3e-51;  
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DTGVESELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGPEGVK 60  
Db 1957 DTGVESELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGPEGVK 2016  
  
QY 61 DAGYVINLSKDTFIKVPFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120  
Db 2017 DAGYVINLSKDTFIKVPFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 2076  
  
QY 121 EHSQSKSDTKDVTATVLDKNNISSKSTTNNPK 154  
Db 2077 EDHSQSKSDTKDVTATVLDKNNISSKSTTNNPK 2110

RESULT 3  
T18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18467

R;Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A;Reference number: Z18937  
A;Accession: T18467  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-558 <LAW>  
A;Cross-references: UNIPROT:O77355; UNIPARC:UPI000017CC2E; EMBL:AL008970; NID:e1407852;  
C;Genetics:  
A;Map position: 3  
A;Introns: 84/1; 160/1  
A;Note: C0465c

Query Match 13.8%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 0.74;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;  
  
QY 27 IVSBEEDFILPVY-----KGELEKGYQFDGWEISGPEGKK----DAGYVINLSKDTFIKVP 77  
Db 60 ILGFEDDILYCYISQLQSKSEK--KADGEEDKYLNAXKGLNLTGFIGNKKSDIPBEL 117  
  
QY 78 PKKI--EEKKEE-----ENKPTFDVSK-KKDNPNQVNHSQLNE-----SHRK 115  
Db 118 LELLNEEKKKEHIAIDLNENK-TNDIKKVKENENINENVNENKDISNKKDKEHVSHQN 176  
  
QY 116 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTTN 150  
Db 177 EHNINNVNLKKEKEYTDIQRDKRHKRSLSQKSDSKYKRPFRNKRKTSIER-SLSNKRIDE 235  
  
QY 151 NPNK 154  
Db 236 KTNK 239

RESULT 4  
JC5497  
Clausstrin - chicken  
N;Alternate names: keratan sulfate proteoglycan  
C;Species: Gallus gallus (chicken)  
C;Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 09-Jul-2004  
C;Accession: JC5497; PC4334; S37561

R;Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A;Title: Clausstrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A;Reference number: JC5497; MUID:94157526; PMID:7906711  
A;Accession: JC5497  
A;Molecule type: mRNA  
A;Residues: 1-1038 <BURL>  
A;Cross-references: UNIPROT:Q90784; UNIPARC:UPI00000FC87F; EMBL:X67778; NID:g406318; PID  
A;Accession: PC4334

A;Molecule type: Protein  
A;Residues: 79-83;299-412;485-502 <BUR2>  
A;Cross-references: UNIPARC:UPI000017BFF3; UNIPARC:UPI000017BFF4; UNIPARC:UPI000017BFF5  
A;Experimental source: brain  
A;Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervous  
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F;267-270/Region: cell attachment (R-G-D) motif  
F;112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

Query Match 13.6%; Score 108.5; DB 2; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 1.9;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;  
  
QY 21 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFBGKODAGYVINLSKDTFIKVPVK- 79  
Db 595 KPETKTIVAEKDV-----TTKEEQLGKSETSEKQASEKQDVKPKVYKEKSVKCEVKA 646  
  
QY 80 KIEEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQRE-----EHSQSKSDTKDV 133  
Db 647 KPBEKCODEKPKPKKEVSKKEKPLI---KKEEKPKGSDIKKEVKKEVKEKKEGAKKEV 702

RESULT 5  
B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C;Accession: B71609  
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Portea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A;Reference number: A71600; MUID:99021743; PMID:9804551  
A;Accession: B71609  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-665 <GAR>  
A;Cross-references: UNIPROT:O96229; UNIPARC:UPI000017B60A; GB:AE001410; NID:AE001362; NID  
A;Experimental source: clone 3D7  
C;Genetics:  
A;Gene: PFB0680w

Query Match 13.1%; Score 105; DB 2; Length 665;  
Best Local Similarity 24.4%; Pred. No. 2.1;  
Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;  
  
QY 1 DTGVESELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEI--SGFE 57  
Db 128 EKNKINKSDLHRQNELNLQSGK-----NEQDI-----NKNEKGKQ-----DISNSAE 170  
  
QY 58 GKQDAGYVINLSKDTFIKVPFKKIEKKKE-----EENKPTFD-----VSKKKNPQ 103  
Db 171 NKQD-----VKEGVKLEELKKKEEKISDDHKVKEENKSDDKHKVKEENKSDDHK 218  
  
QY 104 VNHSQLNESHKEDLQR-EHSQSKSDTKDVTATVLDKNNISSKSTTNNPK 154  
Db 219 VEENKKSDDHKIEEVKVEEHEDEEE-----DKCKEKKSENKNKNDENK 261

RESULT 6  
T18283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C;Species: Dictyostelium discoideum

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18283  
R;Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; HUGH  
Genetics 148, 1117-1125, 1998  
A;Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1  
A;Reference number: Z14684; MUID:98198836; PMID:9539429  
A;Accession: T18283  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-325 <R18>  
A;Cross-references: UNIPROT:O44016; UNIPARC:UPI000007C824; EMBL:U00796; NID:g2702254; PI  
C;Genetics:  
A;Introns: 85/1

Query Match 13.0%; Score 103.5; DB 2; Length 325;  
Best Local Similarity 23.8%; Pred. No. 1.2;  
Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

QY 11 HRVTVTIONGKEMSTIVSEDFILPVYK-GEL--EKGYQFDGWEISGFROK----- 59  
DB 57 HRTITSIKN--RFSVKKIGDEEKLFRISKNGELIVLNELEFDNPHIK--EGKHLRKSQMF 112  
QY 60 ---KDAGYV-----INLSKDTFTKPV-----FKK----- 80  
DB 113 NHIKDSGYATNWEITEIFLESCTLCKRITATQKRNYSYKRNMIINKLPBEEEEEEEEER 172  
QY 81 --IEEKEEENKPTPDVSKKONQOVNHSQLNESHKE-----DLQREHHSQKSDSTK 131  
DB 173 EEEEEQSEVEKPTTISEEEEEETPAVSEEEEEEEEEETPAVSEEEEEEEEEQDEEKEK 232  
QY 132 D 132  
DB 233 D 233

RESULT 7  
G81339  
probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: G81339  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: G81339  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-312 <PAR>  
A;Cross-references: UNIPROT:Q9PPL5; UNIPARC:UPI00000C1CF0; GB:AL111168; NID  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj0692c

Query Match 12.5%; Score 100; DB 2; Length 312;  
Best Local Similarity 25.3%; Pred. No. 2.2;  
Matches 41; Conservative 32; Mismatches 61; Indels 28; Gaps 8;

QY 17 IONGKEMSTI---VSEEDFILPVYK-----GELSKGYQFDGWEISGFEGKKGAGYV 65  
DB 24 INQKPLDDDLRDEITSDDILRRFRPKKTPNKFELDEBVEYSKTKSNLYLKED---L 80  
QY 66 INL---SKDTFTKPVKKIEEKEENKPT---PDVSKKONP-----QVNHSQLNESHK 115  
DB 81 INVLKEEQSLAKLIFSKMKERKEENKTKNKFIFSRKCANBIKNIQTKIQIOTKSNQA 140  
QY 116 EDLQREHHSQKSDSTKDV--TATVLDKNNISSK--STNNPN 153  
DB 141 TTQTKQKKELTNSIEKTKQKTKTKQKPLIIEKKLDVKNQFN 182

RESULT 8

T25911  
hypothetical protein T23B3.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T25911  
R;Maggi, L.; Le, T.  
submitted to the EMBL Data Library, February 1997  
A;Description: The sequence of C. elegans cosmid T23B3.  
A;Reference number: Z20109  
A;Accession: T25911  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-211 <MAG>  
A;Cross-references: UNIPROT:P91488; UNIPARC:UPI0000007D762; EMBL:U08309; PIDN:AAB42334.1  
C;Genetics:  
A;Experimental source: strain Bristol N2; clone T23B3  
A;Gene: CESP.T23B3.5  
A;Map position: 1  
A;Introns: 30/2; 200/3

Query Match 12.4%; Score 99; DB 2; Length 211;  
Best Local Similarity 29.4%; Pred. No. 1.7;  
Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

QY 57 EGKKGAGYVNLSDTKTIKVPFKIKIEKKEENKPTFDVSKKONQOVNH-----SQLNES 112  
DB 69 EGEKKDGEKKSEKKGDKGKBEKKDEBKDGKDEKDEKDEKDEKDEKDEKDEKDE 128  
QY 113 HRKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
DB 129 EKDDKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEK 170

RESULT 9  
A84152  
hypothetical protein BH4017 [imported] - Bacillus halodurans (strain C-125)  
C;Species: Bacillus halodurans  
C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C;Accession: A84152  
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hir  
Nucleic Acids Res. 28, 4317-4331, 2000  
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A;Reference number: A83650; MUID:20512582; PMID:11058132  
A;Accession: A84152  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-614 <STO>  
A;Cross-references: UNIPROT:Q9K5S1; UNIPARC:UPI000000C4396; GB:AP001520; GB:BA0000004; NI  
A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH4017

Query Match 12.2%; Score 97.5; DB 2; Length 614;  
Best Local Similarity 27.3%; Pred. No. 7.1;  
Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;

QY 6 SELKPHRVTVTIQNGKEMSTIVSE-----EDFILPVYKGELEKGYQFDGW----ETSG 55  
DB 386 SEFPPE--TYTLQTAIQM-TPVINEYSPQTEEF-----ARKAQLDQGWADVSKVY 435  
QY 56 FEGKKGAGYVNLSDTKTIKVPFKIKIEKK--EENKPTFDVSKK-----DNPQVNHSQL 109  
DB 436 FAGRNIASQL-----GKIEEKLQDKYNNVTFDFPKKEVVNVQSPIKSTALSAL 482  
QY 110 NESHRKEDLQREHHSQKSDSTKDVATVLD 139  
DB 483 GKVTIGVDLGRKHYKKGEDLERRLSQIE 512

RESULT 10  
S46817  
hypothetical protein YHR080c - yeast (Saccharomyces cerevisiae)  
C;Species: Saccharomyces cerevisiae

C>Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Oct-2004  
C/Accession: S46817  
R:Pavello, T.  
submitted to the EMBL Data Library, June 1994  
A/Description: The sequence of S. cerevisiae cosmid 9205.  
A/Reference number: S46795  
A/Accession: S46817  
A/Molecule type: DNA  
A/Residues: 1-1345 <FAV>  
A/Cross-references: UNIPROT:P38800; UNIPARC:UPI000013B2B1; EMBL:U10556; NID:G500825; PID:  
C/Genetics:  
A/Cross-references: SGD:S0001122  
A/Map position: 8R  
C/Superfamily: uncharacterized conserved protein  
C/Keywords: transmembrane protein

Query Match 12.2%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 17;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

Qy 39 KGELEKGYQFDGWEISGPEGKK-DAGYVNLNLSKDTFIKPVFKKIEEKEENKPTFDYSK 97  
Db 1109 KGAIKRG-----SVGEQKVSVDYMLSELRDII-----SRAKSKPKVKYMK 1149  
Qy 98 KKDNPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLD--KNNISSKSTTN---NP 152  
Db 1150 SHDKHRPFHSHKVE-----QKSSERSKSDNDKDLITLTHLDFVQNNFSSEIIPNNKLLSP 1201  
Qy 153 NK 154  
Db 1202 QK 1203

RESULT 11  
T20410  
hypothetical protein E02A10.2 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T20410  
R:Thomas, K.  
submitted to the EMBL Data Library, October 1996  
A/Reference number: Z19271  
A/Accession: T20410  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-385 <WIL>  
A/Cross-references: UNIPROT:Q93424; UNIPARC:UPI00000835C8; EMBL:Z81053; PIDN:CAB02877.1;  
A/Experimental source: clone E02A10  
C/Genetics:  
A/Gene: CESP:E02A10.2  
A/Map position: 5  
A/Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 12.1%; Score 97; DB 2; Length 385;  
Best Local Similarity 29.6%; Pred. No. 4.6;  
Matches 42; Conservative 24; Mismatches 48; Indels 28; Gaps 7;

Qy 1 DTGVESELKPHRVVTIQ--NGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEG 58  
Db 256 ETDDFCSLQKNVHCTILRIHKEVAEK--NEEDKKEBEPKKEBKEVEKKE---EDE 310  
Qy 59 KKDAGVNLNLSKDTFIKPVFKKIEEKEENKPTFDVSKKDNPNQVNHSQLNESHKEDL 118  
Db 311 KKEE-----EP--KKEBKEBKEEKE--EVEKKE-----BEKKDEBEPKKEE 349

Qy 119 QREHSHQKSDSTKDVATVLDK 140  
Db 350 KKEBKEBKEDEVEEKSEKVEEK 371

RESULT 12  
T14188  
hypothetical protein T28D5.30 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: T14188  
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro  
submitted to the Protein Sequence Database, August 1999  
A/Reference number: Z17931  
A/Accession: T14188  
A/Molecule type: DNA  
A/Residues: 1-988 <BEV>  
A/Cross-references: UNIPROT:Q9STN4; UNIPARC:UPI000000A3618; EMBL:AL109819  
A/Experimental source: cultivar Columbia; BAC clone T28D5  
C/Genetics:  
A/Gene: ATSP:T28D5.30  
A/Map position: 4  
A/Introns: 162/3; 201/3; 416/3; 439/3; 460/3; 482/3; 504/3; 519/3; 534/3; 559/3; 579/3;  
C/Superfamily: Arabidopsis thaliana hypothetical protein T21C14.40

Query Match 12.1%; Score 97; DB 2; Length 988;  
Best Local Similarity 22.1%; Pred. No. 13;  
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;

Qy 19 NGKMSSTIVSEEDFILPVYKGELEKGY-----QFDGWEISGPEGKDGAVVNLNLSKDTFP 73  
Db 383 NGRQNSNVQSSVDEILSYTDTKVPFSGVGLNVSESDIVELVEDDVRSGAGLSPNVQRDN- 441  
Qy 74 IKPVFKKIEEKK-----EENKPTFDVSKKDNPNQVNHSQLNESHKEDLQREE 122  
Db 442 VEPGDDVRSGDMSNPSPSAANNVRHGPAITFDIMESDNPGRDNVAPMEDHIRSEVQLSP 501  
Qy 123 HSQKSDSTKQVT--ATVLDKNNISSKSTTNPNK 154  
Db 502 HVL---GAKDVTDVSDPTDKVGVDVTDASDPT 532

RESULT 13  
D86432  
hypothetical protein T518.14 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C/Accession: D86432  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: D86432  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-540 <STO>  
A/Cross-references: UNIPROT:Q9SA84; UNIPARC:UPI000000ABF35; GB:AE005172; NID:G4587525; PI:  
C/Genetics:  
A/Map position: 1

Query Match 12.1%; Score 96.5; DB 2; Length 540;  
Best Local Similarity 24.8%; Pred. No. 7.3;  
Matches 36; Conservative 25; Mismatches 59; Indels 25; Gaps 5;

Qy 4 EVSELKPHRVVTITQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDGAG 63  
Db 39 BEDESKP-----EGVEKSAFPKESDPPADLKESEKK-----ALSDLKSKLEEA 82  
Qy 64 YVNLNLSKDTFIKPVFKK---IEEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120  
Db 83 IVDN----TLTKTKKSSPMKEKKEEVKPEAEVEKKGE--EAAEKVEBKESEAVVT 136  
Qy 121 EEHSQKSDSTKQVTATVLDKNNISS 145

Db 137 BEAPKAEVTVAVVTEIIPKEVTT 161

RESULT 14  
T37189  
hypothetical protein C02H7.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37189  
R:Reimbac, D.; Minx, M.  
A:Description: The sequence of C. elegans cosmid C02H7.  
A:Reference number: Z20523  
A:Accession: T37189  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-535 <LEI>  
A:Cross-references: UNIPROT:Q17595; UNIPARC:UPI000008019E; EMBL:U49945; PIDN:AAC47924.1  
A:Experimental source: strain Bristol N2; clone C02H7  
C:Genetics:  
A:Gene: CESP:C02H7.1  
A:Map position: X  
A:Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 12.0%; Score 96; DB 2; Length 535;  
Best Local Similarity 21.3%; Pred. No. 7.9;  
Matches 35; Conservative 28; Mismatches 61; Indels 40; Gaps 5;

Qy 1 DTGEVSELKPHRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKCK 60  
Db 77 DDGSLKNVKAAILI-----SGKDAET-----NQLMLGTWATISFNSRN 116

Qy 61 DAGYVINLSKDTFIKPVFKIEBKKEENKPTFVSVKKONPQVNHSQLNSHREKEDLQR 120  
Db 117 GTG-----EKKKKVKKEDKGDDEKST---TKRSSKKEHEEKSEKKSAAEE 166

Qy 121 EHSKQSDSTK-----DVTATVLDKNNISSKSTNNPNK 154  
Db 167 KEKKKKSSSKSRHKSRRSEKSEKSKKKEKSTTDEKPK 210

RESULT 15  
G71609  
hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: G71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Partea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:9021743; PMID:9804551  
A:Accession: G71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2500 <GAR>  
A:Cross-references: UNIPROT:Q96223; UNIPARC:UPI0000078301; GB:AE001408; GB:AE001362; NID  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0650w

Query Match 12.0%; Score 96; DB 2; Length 2500;  
Best Local Similarity 26.1%; Pred. No. 44;  
Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

Qy 9 KPHRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKCKDAGYVNL 68  
Db 2173 KPYKT-----ENNK-----NEGNEILKKYSIENEENKKNYDKQENECILDKTQCNYVT 2223

Qy 69 -----SKDTFIKPVFKIEBKKEENKPTFVSVKKONPQVNHSQLNE-----SHRKED 117  
Db 2224 KEKNLNDKNSPFSNIVKVLBEESKSDKRD-----DKNDNTRKNLNDKNSPFSNIVK 2280

Qy 118 LQREHSQKSDSTKDVATVATVLDKNNISSKSTNNPNK 154  
Db 2281 LEBEE---KSDDKRD-----DKNDNTRKNLNDK 2308

RESULT 16  
T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C:Accession: T47835  
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24475  
A:Accession: T47835  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <NVA>  
A:Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI000009E0AD; EMBL:AL138658  
A:Experimental source: cultivar Columbia; BAC clone T209  
C:Genetics:  
A:Map position: 3  
A:Introns: 158/2; 329/3  
A:Note: T209.90

Query Match 12.0%; Score 95.5; DB 2; Length 644;  
Best Local Similarity 23.4%; Pred. No. 11;  
Matches 34; Conservative 22; Mismatches 58; Indels 31; Gaps 4;

Qy 18 QNGKMSSTIVSE-----DFILPVYKGELEKGYQFDGWEISGFEKCKDAGYVINLSKDTF 73  
Db 519 ENSKTEKKTADVADKKSVADFLKRIKNSPQK-----ETTSKNQKNDGNV----- 565

Qy 74 IKPVFKIEBKKEENKPTFVSVKKONPQVNHSQLNSHREKEDLQREHSQKSDS----- 129  
Db 566 -----KENDHQKSKDGVKNSKVKPRELSSSTGKKVEVENNNSKSSSKKQ 615

Qy 130 TKQVATVLDKNNISSKSTNNPNK 154  
Db 616 TKETAETVATGKRGESGKDKQPRK 640

RESULT 17  
I37271  
cylicin II - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 09-Jul-2004  
C:Accession: I37271; S52774  
R:Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.  
Exp. Cell Res. 218, 174-182, 1995  
A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: th  
A:Reference number: I37271; MUID:95255491; PMID:7737358  
A:Accession: I37271  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-348 <HRS>  
A:Cross-references: UNIPROT:Q14093; UNIPARC:UPI0000128C36; EMBL:Z46788; NID:g758586; PI

Query Match 11.9%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 5.8;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

Qy 39 KGELEKGYQFDGWEISGFEKCKDAGYVINLSKDTFIKPVFKIEBKKEENKPTF-----DV 95  
Db 205 ESEGEKG-----GTEKDSKKGKDS-----KKGKDSAEIQLQAVKDKDEDCKDANKGDE 256

Qy 96 SK--KKONPQVNHSQLN-----ESHRKEDLQREHSQKSDSTKD-----VTATVLDKNNI 143  
Db 257 SKDAKDAKEIKKGKDKKPKSPSTDSKDDVKKE---SKDATKDAKVAKDKTEKESA 313

Qy 144 SSK 146  
Db 314 DSK 316



T05612  
Hypothetical protein F9D16.270 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T05612  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.P.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05612  
A:Molecule type: DNA  
A:Residues: 1-456 <BEV>  
A:Cross-references: UNIPROT:Q9SUP7; UNIPARC:UPI00000A81F5; EMBL:AL035394  
A:Experimental source: cultivar Columbia; BAC clone F9D16  
C:Genetics:  
A:Map position: 4  
A:Introns: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
A:Note: F9D16.270

Query Match 11.7%; Score 93.5; DB 2; Length 456;  
Best Local Similarity 25.0%; Pred. No. 10;  
Matches 39; Conservative 28; Mismatches 58; Indels 31; Gaps 8;

QY 4 EYSELKP-HRVVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWIEISGFGKKDA 62  
DB 249 EKDPLKPKIPVSAFLIVANERRAALREKNKSVVEVAK-----ITGEWKNLSDKKKA 300

QY 63 GY--VINLSKDTFIKPVFKKIEBKEEENKPTFDVSKKKNQPNQVNHSQLNESHRKEDLQR 120  
DB 301 PYEKVAKKNKTYLQ-AMEEYRTKEE-----ALSQKE-----EEELLKLKQEAQLQM 349

QY 121 EHSQKSDSTKDVATVTLDKNNISSKSTNN--PNK 154  
DB 350 LKKKEKNTDN-----LINKERATKKCKNENVDPNK 378

RESULT 23  
B72291  
Hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C:Accession: B72291  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: B72291  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <ARN>  
A:Cross-references: UNIPROT:Q9XOM6; UNIPARC:UPI00000C12ED; GB:AB001771; GB:AB000512; NID  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1142

Query Match 11.6%; Score 93; DB 2; Length 219;  
Best Local Similarity 26.2%; Pred. No. 4.9;  
Matches 42; Conservative 28; Mismatches 50; Indels 40; Gaps 9;

QY 8 LKPHRVVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQ--FDGWIEISG-----PFGK 59  
DB 64 LIPNFVVELYATDKTTLFAKE-----VLGEESVSYRDLFAGFGVGRGTPTPFFPKGK 116

QY 60 KDAGIVIN-LSKDTIKPVFKKIEBKEEENKPTFDVSKKKNQPNQVNHSQLNESHRKED- 117  
DB 117 EGLGLVPGVVDKDNFIK-ILKYVAQLKED-----FQTYLKKDDPPVGBPLIIEIP-KSDA 170

QY 118 ---LQREHHSQKSDS-----TKDVTATVLDK 140  
DB 171 DFVLEKDNVAKVDTVPNEVRDRDRIYVTDSPDVAKTLQEK 210

RESULT 24  
T32879  
Hypothetical protein C17F3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T32879  
R:Gatung, S.; Scheet, P.  
submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of C. elegans cosmid C17F3.  
A:Reference number: Z21240  
A:Accession: T32879  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-253 <GAT>  
A:Cross-references: UNIPROT:O44948; UNIPARC:UPI0000074BB9; EMBL:AF043692; PIDN:AAB97531  
A:Experimental source: strain Bristol N2; clone C17F3  
C:Genetics:  
A:Gene: CESP:C17F3.3  
A:Map position: 1  
A:Introns: 41/1

Query Match 11.4%; Score 91; DB 2; Length 253;  
Best Local Similarity 32.6%; Pred. No. 8.1;  
Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;

QY 76 PVFKKIEBKE---EENKPTFDVSKK-----KDNQVNHSQLNESHR----- 114  
DB 61 FVAPKVEBKEEKEEKKADDEKKTEBKKDKKISVKKTQETKSKDKK 120

QY 115 ---KEDLQREHHSQKSDSTKDVATVTLDKNNISSK 146  
DB 121 DEKDEKESKESKDEKKKDEKDEKKDEKK 155

RESULT 25  
G96542  
Hypothetical protein P17J6.14 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: G96542  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonci  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G96542  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-629 <STO>  
A:Cross-references: UNIPROT:Q9C6P8; UNIPARC:UPI00000A743C; GB:AB005173; NID:g11054631;  
C:Genetics:  
A:Gene: P17J6.14  
A:Map position: 1

Query Match 11.3%; Score 90.5; DB 2; Length 629;  
Best Local Similarity 19.5%; Pred. No. 24;  
Matches 33; Conservative 37; Mismatches 66; Indels 33; Gaps 5;

QY 10 PHRVVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWIEISGFGKKDAGVIVNLS 69  
DB 423 PHLETAKPTKDSANEQITVEADVAMNPIVEKAMSEWVEAGAAINPIVEADG-----A 476

QY 70 KDTFIKPVFKKIEBKEE-----EENKPTFDVSKKKNQ--VNHSQLESHRKE 116  
DB 477 MNPIVEKAMSQLVEADAAINQAVDANFQAPTGNDDAESDDSEFVSHS---ETLNPP 533

QY 117 DLQREHHSQKSDSTKDVATVTLDKNN-----ISSKSTTNNPNK 154







## RESULT 34



Db 257 RYSDVKIPVYDSELTSEPSKNGESNTNEEKEDISTENHLESTALNIQQSDSTPTP 316  
QY 133 -----VTATVLDKNNISSKSTNN 151  
Db 317 MEEDVVTETVKTETSETDMKLLSQN 340

RESULT 40  
T39009  
Hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T39009  
R:Gentles, S.; Church, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: Z21815  
A:Accession: T39009  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1888 <GEN>  
A:Cross-references: UNIPROT:O14207; UNIPARC:UPI000013AA2P; EMBL:Z98531; PIDN:CAB11064.1;  
A:Experimental source: strain 972h-; cosmid c6B12  
C:Genetics:  
A:Gene: SPDB:SPAC6B12.02c  
A:Map position: 1  
C:Superfamily: Schizosaccharomyces hypothetical protein SPAC6B12.02c

Query Match 11.1%; Score 89; DB 2; Length 1888;  
Best Local Similarity 23.9%; Pred. No. 1.1e+02;  
Matches 37; Conservative 29; Mismatches 59; Indels 30; Gaps 7;

QY 24 SSTIVSEEDP-----ILPVY---KGELEKGYQPDGWEISGFE-----GKKDAG 63  
Db 390 SSSLTSENPFQLNVAANAVSTIPVYRTTKTKMKKN-RFKYVEVEKLPDLILES YGKKAPK 448

QY 64 YVINLSKDTFIKPVFKTEEEKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQREEH 123  
Db 449 FLRVFARSSSHIP--KMIRRKQMDSKKYFPDKESDRQVIDQVLSDWYSGKHEL VQOSH 506

QY 124 SQKSDS-TKDVTATVLDKN-----NISKSTNN 151  
Db 507 SYKKPDSKSGVGNIFSVNSKKHSVNINAKTAANN 541

Search completed: April 24, 2006, 15:01:15  
Job time : 18.1924 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 12.0582 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGEVSELKPHRVTVTIQNG.....ATVLDKNNISSKSTNNPNK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:\*

- 1: /SIDSS/ptcdat2/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /SIDSS/ptcdat2/pubpaa/US06\_NEW\_PUB.pep:\*
- 3: /SIDSS/ptcdat2/pubpaa/US07\_NEW\_PUB.pep:\*
- 4: /SIDSS/ptcdat2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 5: /SIDSS/ptcdat2/pubpaa/US09\_NEW\_PUB.pep:\*
- 6: /SIDSS/ptcdat2/pubpaa/US10\_NEW\_PUB.pep:\*
- 7: /SIDSS/ptcdat2/pubpaa/US11\_NEW\_PUB.pep:\*
- 8: /SIDSS/ptcdat2/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	118	14.8	746	6	US-10-793-626-652
2	108	13.5	354	7	US-11-189-817-2
3	106	13.3	651	7	US-11-128-660-1
4	90	11.3	501	6	US-10-485-517-381
5	90	11.3	645	6	US-10-485-517-244
6	90	11.3	886	7	US-11-087-099-11456
7	87	10.9	140	7	US-11-096-568A-4771
8	86	10.8	443	7	US-11-188-298-1015
9	86	10.8	700	7	US-11-196-475-74
10	85	10.6	472	6	US-10-793-626-658
11	85	10.6	700	7	US-11-196-475-66
12	85	10.6	708	7	US-11-196-475-76
13	85	10.6	720	6	US-10-793-626-2058
14	84.5	10.6	734	7	US-11-072-512-2272
15	84.5	10.6	1694	7	US-11-052-554A-83
16	83.5	10.5	493	7	US-11-096-568A-3070
17	83.5	10.5	493	7	US-11-096-568A-3071
18	83.5	10.5	510	7	US-11-096-568A-3069
19	82.5	10.3	305	7	US-11-096-568A-17853
20	82.5	10.3	327	7	US-11-096-568A-17852
21	82.5	10.3	386	7	US-11-096-568A-17851
22	82.5	10.3	1036	7	US-11-096-568A-28315
23	82.5	10.3	1070	7	US-11-087-099-5657
24	82.5	10.3	1070	7	US-11-096-568A-28314
25	82.5	10.3	1276	7	US-11-096-568A-28313

26	82	10.3	439	7	US-11-188-298-15964	Sequence 15964, A
27	82	10.3	439	7	US-11-188-298-16606	Sequence 16606, A
28	82	10.3	943	6	US-10-475-204-34	Sequence 34, Appl
29	81.5	10.2	313	7	US-11-096-568A-32043	Sequence 32043, A
30	81.5	10.2	425	7	US-11-096-568A-32042	Sequence 32042, A
31	81.5	10.2	434	7	US-11-096-568A-32041	Sequence 32041, A
32	81.5	10.2	8746	7	US-11-098-686-10232	Sequence 10232, A
33	81	10.1	299	6	US-10-793-626-1888	Sequence 1888, Ap
34	81	10.1	785	6	US-10-793-626-464	Sequence 464, App
35	80.5	10.1	248	6	US-10-793-626-464	Sequence 464, App
36	80	10.0	244	7	US-11-096-568A-11931	Sequence 11931, A
37	79	9.9	339	7	US-11-096-568A-4567	Sequence 4567, Ap
38	79	9.9	568	6	US-10-793-626-2482	Sequence 2482, Ap
39	79	9.9	693	7	US-11-196-475-68	Sequence 68, Appl
40	79	9.9	732	7	US-11-124-368A-248	Sequence 248, Appl
41	79	9.9	803	7	US-11-124-368A-241	Sequence 241, App
42	79	9.9	803	7	US-11-124-368A-242	Sequence 242, App
43	79	9.9	1145	6	US-10-793-626-1432	Sequence 1432, Ap
44	78.5	9.8	258	7	US-11-098-686-10475	Sequence 10475, A
45	78.5	9.8	635	6	US-10-821-234-1573	Sequence 1573, Ap

ALIGNMENTS

RESULT 1

US-10-793-626-652  
; Sequence 652, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 652  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-652

Query Match	14.8%	Score 118;	DB 6;	Length 746;
Best Local Similarity	27.2%	Pred. No. 0.026;		
Matches	47;	Conservative	23;	Mismatches 57; Indels 46; Gaps 9;
Qy	1	DTGEVSELKPHRVTVTIQNGKMSSTVSEBDFILPVYK-----GELKGYQPDGM--- 51		
Db	586	DSVNAQSLKP-----ITINGKQIKQKQSVKSTKVLPHSKVMLMTDGLTMP-DMTGWTKE 640		
Qy	52	BISGFE-----GKDDAGYVIN--LSKDTPIKPFVKIKRKEENKPTDVS-----KK 98		
Db	641	DVLAPEDLTKLVKSTKNGFVTVNQSIKQIIL-----NKKLEVLSAEDT 687		
Qy	99	KDPQVNHSQLNEHRKEDLQREHSHKSGKSDTKDTATVLDKNNISSKSTNN 151		
Db	688	DDQEKTDSDSKSKKDAEDHSNTSSSTKN-----DKSNADSKNDSDD 734		

RESULT 2

US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US20060030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRULLHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM

```
; TITLE OF INVENTION: MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES
; FILE REFERENCE: 275601US0
; CURRENT APPLICATION NUMBER: US/11/189,817
; CURRENT FILING DATE: 2005-07-27
; PRIOR APPLICATION NUMBER: 60/598,062
; PRIOR FILING DATE: 2004-08-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-189-817-2

Query Match          13.5%; Score 108; DB 7; Length 354;
Best Local Similarity 23.2%; Pred. No. 0.073;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 17 IQNGKEMSTIVSEBDFILPVYKGLKGYQPDGWEISGF--EGKKDAG-----YVINLS 69
DB 181 VLKAKASS-----YDYL-----GWFGGVPEHKKEENMLSHLYVSSKD 221
QY 70 KDTFKPVFKIIEKKEE-----ENKPTDVSCKKONPVQNHSQLNESHKRE 116
DB 222 KENISKENDVDLDEKEBAEETEEBELEKNEETESEISEDEEBEERKEEBENDKK 281
QY 117 DLORREHSKOSTKDVATVLDKKNISKSSTNN 151
DB 282 EQEKEQSNENNDDQKDMEA-----QNLISKQNNN 311

RESULT 3
US-11-128-660-1
; Sequence 1, Application US/11128660
; Publication NO. US20060024324A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plasmo
; TITLE OF INVENTION: falciparum
; FILE REFERENCE: 15007dk
; CURRENT APPLICATION NUMBER: US/11/128,660
; CURRENT FILING DATE: 2005-05-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-128-660-1

Query Match          13.3%; Score 106; DB 7; Length 651;
Best Local Similarity 22.8%; Pred. No. 0.23;
Matches 43; Conservative 35; Mismatches 55; Indels 56; Gaps 9;

QY 7 ELKPRVTVTIQNGKEM-----SSTIVSEBDFILPVYKGLK----- 44
DB 432 ETEHEETVQSNEPEKADNGVSNQNNELNEFV-----ESEKSEHARSKAKEA 485
QY 45 -GYQPD-GWEISGF--EGKKDAG-----YVINLSKDTFKIPVKIIEBKKEENKPTFDV 95
DB 486 SSYDYLGWFGGVPEHKKEENMLSHLYVSSKKNISKENDDVDLDE-KEEAESETREE 544
QY 96 SKKONPVQNHSQLN-----BSHRKEDLQREHSQKSDSTKDVATVLDKNN 142
DB 545 ELEEKNEETESEISEDEEBEERKEEBEENKKEQKEQSNENNDDQKDMEA-----QN 599
QY 143 ISSKSTNN 151
DB 600 LISKQNNN 608

RESULT 4
US-10-485-517-381
```

```
; Sequence 381, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-381

Query Match          11.3%; Score 90; DB 6; Length 501;
Best Local Similarity 21.4%; Pred. No. 3.7;
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;

QY 4 EVSELKPRVTV--TIQNGKEMSTIVSEBDFILPVYKGELE-KGYQF-----DGWEIS 54
DB 197 KMTDLQDTKYVYVESVENNESMMDTFVKH-----PIKTGMLNGKKYMYMVTNDDYWKDF 251
QY 55 GPEGKK-----DAGYVINL-SKDTPI 74
DB 252 MVEQVRVTTISKDANNTRTIIPFYEGKTYDAIVKVVHTYIDYGGYHVRIVDKEAPT 311
QY 75 KPVFKIIEKKEENKPTFDV-----SKCKONPVQNHSQLNESHKEDLQ-----REEHSQ 125
DB 312 KANTDKSNKKEQDQNSAKKEATPAT?SKPTSPVKESESQKDSQKDDNKQLPSVEKENDA 371
QY 126 KSDSTKQVT-ATVLDKKNISKSSTNNPNK 154
DB 372 SSESQKDKTPATKPTKGEVESSTT--PTK 399

RESULT 5
US-10-485-517-244
; Sequence 244, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 244
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-244

Query Match          11.3%; Score 90; DB 6; Length 645;
Best Local Similarity 21.4%; Pred. No. 5;
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;
```







```
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-76

Query Match
Best Local Similarity 10.6%; Score 85; DB 7; Length 708;
Matches 43; Conservative 22; Mismatches 53; Indels 38; Gaps 8;

QY 13 VTTVTQNGKMSSTTVSEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVIMLSK-- 70
D  :|||: :
Db 258 ITETIENLRDQLEKATDSE-----HKXIES--QVDAKKCKEKLKKKA---INLDKRAQ 306
QY 71 -----DTPFKVPKKEEKEENKP-TDVSCKKONPQVNSHLSHRSK 115
D  :|||: :
Db 307 QKLSAENLDVQRDTRVKIQEDINEINKEKNLPKPGDVSPKVDKQL---QIKES--L 361
QY 116 EDLQRE-----EHSQKSDTKDVTATVLDKNNISK 146
D  :|||: :
Db 362 EDLQRLKAGDNGKRIEIKRIKRDDELLKSK 397

RESULT 13
US-10-793-626-2058
; Sequence 2058, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2058

Query Match
Best Local Similarity 10.6%; Score 85; DB 6; Length 720;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 81 IEKKEEENKPTFDVSKKD-----NPQVNSHLSHRSKEDLQREHSQKSDTK----D 132
D  :|||: :
Db 38 LEEEQIKALDKKFKASQAKDTNKQNTNNHOKSNKQNSNDKKEQKQSKNSKPTKKKEQN 97
QY 133 VTATVLDKNNISSKSTTNPNK 154
D  :|||: :
Db 98 NKGQKQNNKNTKRNKRNK 119

RESULT 14
US-11-072-512-2272
; Sequence 2272, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
```

```
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SERI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2272
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2272

Query Match
Best Local Similarity 10.6%; Score 84.5; DB 7; Length 734;
Matches 39; Conservative 41; Mismatches 58; Indels 41; Gaps 11;

QY 1 DTG-EVSELKPHRVTV-----TIQNGKMSSTIVSE--DFILPVYKG 40
D  :|||: :
Db 169 DTGIVSEVDPSIAAKDGRIGRILQINGEDVQNRREAVALLSNDECKRIVLLVARP 228
QY 41 ELEKGYQFDGWEISGFGKKDAGVIMLSKDTF-----IKPVFKKIEE-KKEEENKPT 92
D  :|||: :
Db 229 EIQLD---EGW-----LEDERNE-FLEELINLEMLEEENEAQPTANEVQPKQEEBEGT 280
QY 93 FDSVSKKONPQVNSHLSHRSKEDLQREHSQKSDTKDVTATVL-DKNNISKSTT 149
D  :|||: :
Db 281 TDTATSSN---NHEKDSGVGRTDESLRNDESSQENAAEDPNSTLSKSKRDLGQSOT 336

RESULT 15
US-11-052-554A-83
; Sequence 83, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 83
; LENGTH: 1694
; TYPE: PRT
; ORGANISM: Haemophilus influenzae Rd
US-11-052-554A-83

Query Match
Best Local Similarity 10.6%; Score 84.5; DB 7; Length 1694;
Matches 24; Conservative 15; Mismatches 15; Indels 3; Gaps 1;

QY 66 INLSKDTFFIKPVFKKIEEKEENKPTFDVSKKONPQVNSHLSHRSKEDLQREHS- 124
D  :|||: :
Db 1288 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSDPKSRRSI 1347
QY 125 --QKSDSTKDVATVLDKNNISSKSTTNPNK 154
D  :|||: :
```

```
Db      1348 SQPQTSABETTAASTDETTIADNSKRSKPNR 1379

RESULT 16
US-11-096-568A-3070
; Sequence 3070, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 2005-04-01
; SEQ ID NO 3070
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 15172486
US-11-096-568A-3070

Query Match      10.5%; Score 83.5; DB 7; Length 493;
Best Local Similarity 25.7%; Pred. No. 13;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

Qy      16 TIONGKEM--SSTIVSE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 63
Db      80 TQSSGKQTADANTIVTEKKPGKVPVKKKIKTPVSK---KKDETADSNKTETLSKKDDEG 136

Qy      64 YVI-----NLSKDTFIKPVFKKIE--KKEENKPTFDVSKKKNPQVNH 107
Db      137 NVAVQAQDDTOSTGKTANADTTVTPEVKTKGVVPPKQSKTPT---SEKRDN--TADS 191

Qy      108 QLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 154
Db      192 SKTETKSDKDDKEER-----VTGEKSGAKTDKLKASDKDVTNVKKG 233

RESULT 17
US-11-096-568A-3071
; Sequence 3071, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 2005-04-01
; SEQ ID NO 3071
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 16625552
US-11-096-568A-3071

Query Match      10.5%; Score 83.5; DB 7; Length 493;
Best Local Similarity 25.7%; Pred. No. 13;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

Qy      16 TIONGKEM--SSTIVSE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 63
Db      80 TQSSGKQTADANTIVTEKKPGKVPVKKKIKTPVSK---KKDETADSNKTETLSKKDDEG 136

Qy      64 YVI-----NLSKDTFIKPVFKKIE--KKEENKPTFDVSKKKNPQVNH 107
```

```
Db      137 NVAVQAQDDTOSTGKTANADTTVTPEVKTKGVVPPKQSKTPT---SEKRDN--TADS 191
Qy      108 QLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 154
Db      192 SKTETKSDKDDKEER-----VTGEKSGAKTDKLKASDKDVTNVKKG 233

RESULT 18
US-11-096-568A-3069
; Sequence 3069, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; NUMBER OF SEQ ID NOS: 2005-04-01
; SEQ ID NO 3069
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Ceres Seq. ID no. 15172485
US-11-096-568A-3069

Query Match      10.5%; Score 83.5; DB 7; Length 510;
Best Local Similarity 25.7%; Pred. No. 13;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

Qy      16 TIONGKEM--SSTIVSE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 63
Db      97 TQSSGKQTADANTIVTEKKPGKVPVKKKIKTPVSK---KKDETADSNKTETLSKKDDEG 153

Qy      64 YVI-----NLSKDTFIKPVFKKIE--KKEENKPTFDVSKKKNPQVNH 107
Db      154 NVAVQAQDDTOSTGKTANADTTVTPEVKTKGVVPPKQSKTPT---SEKRDN--TADS 208

Qy      108 QLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 154
Db      209 SKTETKSDKDDKEER-----VTGEKSGAKTDKLKASDKDVTNVKKG 250

RESULT 19
US-11-096-568A-17853
; Sequence 17853, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; SEQ ID NO 17853
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(305)
; OTHER INFORMATION: Ceres Seq. ID no. 12361233
US-11-096-568A-17853

Query Match      10.3%; Score 82.5; DB 7; Length 305;
Best Local Similarity 19.6%; Pred. No. 8.7;
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;
```

QY 10 PHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVYNLS 69  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 40 PDLVDCSLNGDAGSKKKAESKSRPVAK-----ETPSLEDSNEKKKTQKAS 87  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 70 KDTFIKVPFKLIEKKKEENKP-----TFDVSKKDNQVNHSQLNESHKEDL 118  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 88 NQHSVK---KDIEESNESVKPQVGSYSPYGFKCDERAKRRBFYSKLEEKIHAQEL 144  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 119 QREHHSQKSDSTKDVATVLDKN-----ETPSLEDSNEKKKTQKAS 141  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 145 EKSNIQAKSKETEAEELKMLRSLNFKATPMPSPYKEPPPPKVELKKIPTTTRARSPKCLR 204  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 142 --NISKSTTNPN 153  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 205 SKNTSSGGTEGNPN 218  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :

## RESULT 20

US-11-096-568A-17852  
; Sequence 17852, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 17852  
; LENGTH: 327  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(327)  
; OTHER INFORMATION: Ceres Seq. ID no. 12361232  
US-11-096-568A-17852

Query Match 10.3%; Score 82.5; DB 7; Length 327;  
Best Local Similarity 19.6%; Pred. No. 9.5;  
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;  
QY 10 PHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVYNLS 69  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 62 PDLVDCSLNGDAGSKKKAESKSRPVAK-----ETPSLEDSNEKKKTQKAS 109  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 70 KDTFIKVPFKLIEKKKEENKP-----TFDVSKKDNQVNHSQLNESHKEDL 118  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 110 NQHSVK---KDIEESNESVKPQVGSYSPYGFKCDERAKRRBFYSKLEEKIHAQEL 166  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 119 QREHHSQKSDSTKDVATVLDKN-----ETPSLEDSNEKKKTQKAS 141  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 167 EKSNIQAKSKETEAEELKMLRSLNFKATPMPSPYKEPPPPKVELKKIPTTTRARSPKCLR 226  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 142 --NISKSTTNPN 153  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 227 SKNTSSGGTEGNPN 240  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :

## RESULT 21

US-11-096-568A-17851  
; Sequence 17851, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 17851

; LENGTH: 386  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(386)  
; OTHER INFORMATION: Ceres Seq. ID no. 12361231  
US-11-096-568A-17851

Query Match 10.3%; Score 82.5; DB 7; Length 386;  
Best Local Similarity 19.6%; Pred. No. 12;  
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;  
QY 10 PHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVYNLS 69  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 121 PDLVDCSLNGDAGSKKKAESKSRPVAK-----ETPSLEDSNEKKKTQKAS 168  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 70 KDTFIKVPFKLIEKKKEENKP-----TFDVSKKDNQVNHSQLNESHKEDL 118  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 169 NQHSVK---KDIEESNESVKPQVGSYSPYGFKCDERAKRRBFYSKLEEKIHAQEL 225  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 119 QREHHSQKSDSTKDVATVLDKN-----ETPSLEDSNEKKKTQKAS 141  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 226 EKSNIQAKSKETEAEELKMLRSLNFKATPMPSPYKEPPPPKVELKKIPTTTRARSPKCLR 285  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 142 --NISKSTTNPN 153  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 286 SKNTSSGGTEGNPN 299  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :

## RESULT 22

US-11-096-568A-28315  
; Sequence 28315, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28315  
; LENGTH: 1036  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(1036)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712010  
US-11-096-568A-28315

Query Match 10.3%; Score 82.5; DB 7; Length 1036;  
Best Local Similarity 24.1%; Pred. No. 38;  
Matches 38; Conservative 29; Mismatches 68; Indels 23; Gaps 6;  
QY 1 DTGEVS-----ELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEI 53  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 504 ETGDTSDPSAKANEQTPAKTIVKKKIIKRVAKRVAIDNKM---DGDGSKDSDSEKVV 560  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 54 SGFEGKK--DAGYVINLSKDTPIKVPFKLIEKKKEENKPTFDVSKKKNQVNHSQLNE 111  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 561 MEV-GKKSSDSGSV-----EMKPTAESLESDVDENASKTVDVKQETGSPDTKKKEGAS 612  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
QY 112 SHRKEDLQREH---SOKSDSTKDVATVLDKNISSK 146  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
Db 613 SSSKQTKTGTGDKKAERKQNSSETWSBGKKIDRNNTDEK 650  
| | | | | : | | | | : | | | | : | | | | : | | | | : | | | | : | | | | :  
RESULT 23  
US-11-087-099-5657  
; Sequence 5657, Application US/11087099  
; Publication No. US20060041961A1

```
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 5657
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Bacillus anthracis str. Ames
US-11-087-099-5657

Query Match          10.3%; Score 82.5; DB 7; Length 1070;
Best Local Similarity 24.1%; Pred. No. 40;
Matches 37; Conservative 24; Mismatches 66; Indels 27; Gaps 6;

Qy  2 TGEVSELKP-----HRVTVTIQNGKMSSTIVSEEDFILPVY-----KGELEKGY 46
Db  411 TNEIRDRIQELGKRIKIDAQROKQVFLDEAYMNEEVKIPYDVNGTALQNIWKSEGS 470
Qy  47 QPDG---WEISG---FEGKADAGYVINLSKDTFIKPVFKKIEEKKEENKPTFDVSKKK- 99
Db  471 ITNGVIKWSLGERKMYEPMADGE-----SKIRFQGRVIGNIVEKREESSNVIQMKLQY 526
Qy  100 -DNPQVNSQLNESHRKEDLQREHSQKSDSTKD 132
Db  527 MNKHNFERNVNTPTIKGDLTLTKALKITDGKKE 560

RESULT 24
US-11-096-568A-28314
; Sequence 28314, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28314
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc feature
; LOCATION: (1)-(1070)
; OTHER INFORMATION: Ceres Seq. ID no. 2712009
US-11-096-568A-28314

Query Match          10.3%; Score 82.5; DB 7; Length 1070;
Best Local Similarity 24.1%; Pred. No. 40;
Matches 38; Conservative 29; Mismatches 68; Indels 23; Gaps 6;

Qy  1 DTGEVS-----ELKPHRVVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 53
Db  538 ETGDTSDPSAKANEQTAKTIKVKKIIKRVAKRVAEIDNKM---DGDSSKKGDSDEKKV 594
Qy  54 SGFEGKK--DAGYVINLSKDTFIKPVFKKIEEKKEENKPTFDVSKKKDNQVNSQLNE 111
Db  595 MEV-GKKSDSGSV-----EMKPTAESLEDVKDENASKTVDVKQGTSPDTKKKGAS 646
Qy  112 SHRKEDLQREH---SQKSDSTKDVATATVLDKNNISSK 146
Db  647 SSSKDKTKTGEDKKAEEKNNSETMSEGGKIDRNNNTDEK 684

RESULT 25
US-11-096-568A-28313
; Sequence 28313, Application US/11096568A
; Publication No. US20060048240A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28313
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc feature
; LOCATION: (1)-(1276)
; OTHER INFORMATION: Ceres Seq. ID no. 2712008
US-11-096-568A-28313

Query Match          10.3%; Score 82.5; DB 7; Length 1276;
Best Local Similarity 24.1%; Pred. No. 49;
Matches 38; Conservative 29; Mismatches 68; Indels 23; Gaps 6;

Qy  1 DTGEVS-----ELKPHRVVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 53
Db  744 ETGDTSDPSAKANEQTAKTIKVKKIIKRVAKRVAEIDNKM---DGDSSKKGDSDEKKV 800
Qy  54 SGFEGKK--DAGYVINLSKDTFIKPVFKKIEEKKEENKPTFDVSKKKDNQVNSQLNE 111
Db  801 MEV-GKKSDSGSV-----EMKPTAESLEDVKDENASKTVDVKQGTSPDTKKKGAS 852
Qy  112 SHRKEDLQREH---SQKSDSTKDVATATVLDKNNISSK 146
Db  853 SSSKDKTKTGEDKKAEEKNNSETMSEGGKIDRNNNTDEK 890

RESULT 26
US-11-188-298-15964
; Sequence 15964, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 15964
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus DSM 3638
US-11-188-298-15964

Query Match          10.3%; Score 82; DB 7; Length 439;
Best Local Similarity 18.1%; Pred. No. 15;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

Qy  45 GYQPDGWEISGFEGKKDAGYVINLSKDTFI-----KP-----VFK 79
Db  45 GISFDGSSVPGFGQIEDSLVFKAADPTTYVEVPMDNVARVYGYKDNKPYGADPRGILK 104
Qy  80 KIEEKKEEN-----KPTFDVSKKKDN----- 101
Db  105 RALEBELEKGYKAYIGPEPEFYLFKKGWTWELEIPDVGGYFDILTLDKARDIRREIAEYM 164
Qy  102 -----PQVNSQLNESHRKEDLQREHSQKSD---STKDVATATVLDKNNI 143
Db  165 PSFGLIPEVLHHEVGKQAHEIDFRYDEALKATADNIVSFYKTKAVAEHGL 215

RESULT 27
US-11-188-298-16606
```



Query Match	10.2%	Score 81.5	DB 7	Length 8746
-------------	-------	------------	------	-------------

```

RESULT 34
US-10-793-626-264
; Sequence 264, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 264
; LENGTH: 785
; TYPE: PRT

```

```

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-264

Query Match      10.1%; Score 81; DB 6; Length 785;
Best Local Similarity 24.7%; Pred. No. 37;
Matches 37; Conservative 27; Mismatches 66; Indels 20; Gaps 7;

QY 14 TVTIOGKEMS-STIVSEDFILPVYKGELEKGYQFDGWEISGFGKKGAGY-VINLSKD 71
Db 638 TITERNGNIVCTITVGEEDLMVTNAGVI-----IRLDVHDSQ-NGRAAQGVRLMKUGDG 693
QY 72 TFIKVPFKKIEBKKEEN-----KPTFDVSKKKONPQVNHSQLNESHKEDLQREE 122
Db 694 QFVSTVAKVNEEDNEENADEAQSTTTTETADEVVVD----DQTGNAIHTEGDAEMES 749
QY 123 -HSQKSDSTKYDTATVLDKNNISSKSTNN 151
Db 750 VEPFENDRIDIRQDFMDRVNEDIESADN 779

RESULT 35
US-10-793-626-464
; Sequence 464, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034800S
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 464
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-464

Query Match      10.1%; Score 80.5; DB 6; Length 248;
Best Local Similarity 24.7%; Pred. No. 10;
Matches 40; Conservative 18; Mismatches 63; Indels 41; Gaps 6;

QY 1 DTGEVSELKPHRVTVTIOGKEMS-----STIVSEDFILPVYKGELEKGYQF 48
Db 53 DTNEVHLKEDYLK-TVEN-KKESYDLKGFVDLCNRSIKONEDIL--DYTKLFEARTE 108
QY 49 DGWEISGFGKKGAGYVINLSKDTFIKVPFKKIEBKKEENKPTFDVSKKKONPQVNHQS 108
Db 109 VESDINKAQNKEDA-----SQLKSLKENNQKDTAKK-----Y 143
QY 109 LNESHRKEDLQREHSQKSDSTKYDTATVLDKNNISSKSTNN 150
Db 144 LNSNNDSDSAKEATKNHISPLIDKQITDINKTNISDNHVDN 185

RESULT 36
US-11-096-568A-11931
; Sequence 11931, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
```

```

; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 11931
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(244)
; OTHER INFORMATION: Ceres Seq. ID no. 13659132
US-11-096-568A-11931

Query Match      10.0%; Score 80; DB 7; Length 244;
Best Local Similarity 23.7%; Pred. No. 11;
Matches 40; Conservative 31; Mismatches 56; Indels 42; Gaps 7;

QY 9 KPHRVTVTIOGKEMSSTI-----VSEEDFILPVYKGELEKGYQFDGW---EI 53
Db 95 RPVAVASSLRNKKVSSLVDRKWAAKGELRDEEEEDALEYLERKKRKIDGWRKQOI 154
QY 54 SGFEGKKGAGYVINLSKDTFIKVPFKKIEBKKEENKPTFDVSKKKONPQVNHSQLNESH 113
Db 155 ASGEAKENANFV-PLGGDWRDRVKKRAEAKKEATBPIRAAA-----EQH 199
QY 114 RKG-DLQREHSQKS-----DSTKQVTVLVLDKNNISSKSTNNPNK 154
Db 200 KGEPLDSELSKGFPSGQAYIDESTKQ----VYTGNNLTSETTWDPRPSK 244

RESULT 37
US-11-096-568A-4567
; Sequence 4567, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4567
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(339)
; OTHER INFORMATION: Ceres Seq. ID no. 15219423
US-11-096-568A-4567

Query Match      9.9%; Score 79; DB 7; Length 339;
Best Local Similarity 25.6%; Pred. No. 20;
Matches 42; Conservative 32; Mismatches 54; Indels 36; Gaps 9;

QY 2 TGEVSELKPHRVTVTIOGKEMSSTIVSE-----EDFILPVYK---GELEKG--Y 46
Db 117 TGDVE--RPRV-YDRRSVYDNRSGTGRSNDVKREGGGRGNWGTPEDDIQPVTEPTTEVKS 173
QY 47 QFDGWEISGFGKKGAGYVIN--LSKDTFIKVPFKKIEBKKE-----ENK---PTFD 94
Db 174 EKEGEDATTDATKKEAPEVEQEPEDKEMTLBEYEKILEBKKKALQATKVEERKVDTRK 233
QY 95 VSKKKONPQVNHSQL-----NESHKEDLQREHSQKSDSTKD 132
Db 234 SMQQLSNKKTNDSEIFIKLGSDDKDKRAEKEERAKKSLSLNE 277

RESULT 38
US-10-793-626-2482
; Sequence 2482, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
```





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 77.1723 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773  
Perfect score: 799  
Sequence: 1 DTGEVSELKPHRVTVTIQNG.....ATVLDKNNISSKSTTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main: \*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	799	100.0	773	4	US-10-067-385-8
2	799	100.0	2119	3	US-09-769-744A-28
3	799	100.0	2140	4	US-10-282-122A-73670
4	799	100.0	2140	5	US-10-472-928-1180
5	796	99.6	637	5	US-10-617-320-3169
6	615	77.0	117	3	US-09-765-272-68
7	615	77.0	117	6	US-11-106-649-68
8	118	14.8	778	4	US-10-724-972A-5663
9	110.5	13.8	775	5	US-10-282-122A-70721
10	108	13.5	188	5	US-10-691-672A-7
11	107.5	13.5	470	5	US-10-739-930-6262
12	106	13.3	647	5	US-10-691-672A-3
13	105	13.1	665	3	US-09-820-843A-107
14	103.5	13.0	169	5	US-10-691-672A-2
15	103	12.9	707	4	US-10-282-122A-52942
16	101.5	12.7	564	6	US-11-097-143-13723
17	100	12.5	973	5	US-10-732-923-18783
18	97.5	12.2	1184	4	US-10-282-122A-53254
19	97	12.1	3127	5	US-10-732-923-22588
20	96.5	12.1	540	5	US-10-732-923-22820
21	96.5	12.1	2060	4	US-10-381-596A-2
22	95.5	12.0	1373	5	US-10-732-923-16976
23	95	11.9	948	5	US-10-732-923-4286
24	93.5	11.7	898	4	US-10-425-115-205148
25	93	11.6	869	4	US-10-437-963-12282
26	93	11.6	1529	5	US-10-732-923-8762
27	92.5	11.6	903	4	US-10-282-122A-52328

28	92.5	11.6	2468	4	US-10-755-889-615	Sequence 615, App
29	92.5	11.6	2468	5	US-10-489-740-216	Sequence 216, App
30	92.5	11.6	2519	5	US-10-450-763-46995	Sequence 46995, A
31	92	11.5	1005	4	US-10-437-963-187665	Sequence 187665,
32	91.5	11.5	3124	5	US-10-732-923-22709	Sequence 22709, A
33	90.5	11.3	743	4	US-10-171-311-188	Sequence 188, App
34	90.5	11.3	758	3	US-09-925-299-859	Sequence 859, App
35	90.5	11.3	758	3	US-09-925-299-859	Sequence 859, App
36	90.5	11.3	932	4	US-10-282-122A-52510	Sequence 52510, A
37	90	11.3	442	4	US-10-724-972A-70294	Sequence 5858, Ap
38	90	11.3	645	4	US-10-282-122A-70294	Sequence 70294, A
39	90	11.3	645	5	US-10-470-048B-414	Sequence 414, App
40	90	11.3	654	4	US-10-172-502-10	Sequence 10, Appl
41	90	11.3	654	6	US-11-020-509-10	Sequence 10, Appl
42	90	11.3	1183	6	US-11-097-143-3099	Sequence 3099, Ap
43	89.5	11.2	402	3	US-09-827-664-4	Sequence 4, Appli
44	89.5	11.2	511	4	US-10-289-762-509	Sequence 509, App
45	89.5	11.2	529	3	US-09-827-664-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 799;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 4.1e-61;		
Matches 154;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGFEGKK	60	
Db	620	DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGFEGKK	679	
Qy	61	DAGTVINLSKDTFKIPVFKKIEKKKEENKPTFDVSKKONPQVNHSLNESHRKEDLQR	120	
Db	680	DAGTVINLSKDTFKIPVFKKIEKKKEENKPTFDVSKKONPQVNHSLNESHRKEDLQR	739	
Qy	121	BEHSQSDSTKDVATATVLDKNNISSKSTTNNPNK	154	
Db	740	BEHSQSDSTKDVATATVLDKNNISSKSTTNNPNK	773	

RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins

```
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 799; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 1.4e-60;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFECKK 60
Db 1932 DTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFECKK 1991

QY 61 DAGYVINLSKDTPIKPVFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120
Db 1992 DAGYVINLSKDTPIKPVFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 2051

QY 121 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
Db 2052 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2085

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09

; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match      100.0%; Score 799; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.4e-60;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFECKK 60
Db 1953 DTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFECKK 2012

QY 61 DAGYVINLSKDTPIKPVFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120
Db 2013 DAGYVINLSKDTPIKPVFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 2072

QY 121 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
Db 2073 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2106

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/10472923
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
US-10-472-928-1180

Query Match      100.0%; Score 799; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.4e-60;
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFECKK 60
Db 1953 DTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFECKK 2012

QY 61 DAGYVINLSKDTPIKPVFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120
Db 2013 DAGYVINLSKDTPIKPVFKKIEKKKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 2072

QY 121 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
Db 2073 EEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2106

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```

```
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC THERAPEUTICS
;
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEtical: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-10-617-320-3169

Query Match 99.6%; Score 796; DB 5; Length 637;
Best Local Similarity 99.4%; Pred. No. 5,9e-61;
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGVSSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISFGEGKK 60
DB 450 DTGVSSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISFGEGKK 509
QY 61 DAGYVNLKSDTFIKPVFKIEEKEENKPTFDYKKNQVNHSQLNESHKEDLQR 120
DB 510 DAGYVNLKSDTFIKPVFKIEEKEENKPTFDYKKNQVNHSQLNESHKEDLQR 569
QY 121 EBHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
DB 570 EBHSQKSDSTKDVATVLDKNNISSKSTNNPNK 603

RESULT 6
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US2002006154A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; FILE REFERENCE: PB3402C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

```
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match 77.0%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.8e-46;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 YKGELEKGYQFDGWEISFGEGKKDAGYVNLKSDTFIKPVFKIEEKEENKPTFDVSK 97
DB 1 YKGELEKGYQFDGWEISFGEGKKDAGYVNLKSDTFIKPVFKIEEKEENKPTFDVSK 60
QY 98 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
DB 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 7
US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; FILE REFERENCE: PB3402C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

```
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      77.0%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.8e-46;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 YGELEKGYQPGWEISGEGKDDAGYVNLKDTFKPVPFKIEEKEEENKPTFDVSK 97
Db 1 YGELEKGYQPGWEISGEGKDDAGYVNLKDTFKPVPFKIEEKEEENKPTFDVSK 60

Qy 98 KKDQPNVHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKTTNNPNK 154
Db 61 KKDQPNVHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKTTNNPNK 117

RESULT 8
US-10-724-972A-5663
; Sequence 5663, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: PAT03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; PRIOR FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 5663
; LENGTH: 778
; TYPE: PRT
; ORGANISM: S. epidermidis
US-10-724-972A-5663

Query Match      14.8%; Score 118; DB 4; Length 778;
Best Local Similarity 27.2%; Pred. No. 0.13;
Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;

Qy 1 DTGVESELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQPDGM--- 51
Db 618 DSVNAQSLAP-----ITIGNGKQIQQSVKSGTKVLPKSKVLMVTDGELTMP-DMTGWTKE 672

Qy 52 BISGFE-----GKDDAGYVIN--LSKDTFKPVPFKIEEKEEENKPTFDVS-----KK 98
Db 673 DVLAPEDLTKIKVSTKNGFVTNQISKGQIILK-----NKKIEVLSABDT 719

Qy 99 KKDQPNVHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKTTNN 151
Db 720 DDDQEKTDSDSDNKSKKDQADEHSNTSSSTKN-----DKSNADSKNDSDD 766

RESULT 9
US-10-282-122A-70721
; Sequence 70721, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70721
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70721

Query Match      13.8%; Score 110.5; DB 4; Length 775;
Best Local Similarity 25.0%; Pred. No. 0.6;
Matches 42; Conservative 27; Mismatches 66; Indels 33; Gaps 8;

Qy 1 DTGVESELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQPDGM--- 51
Db 615 DSVNAQSLAP-----ITIGNGKQIQQSVKSGTKVLPKSKVLMVTDGELTMP-DMTGWTKE 669

Qy 52 BISGFE-----GKDDAGYVIN--LSKDTFKPVPFK-----KIEEKEEENKPTFD 94
Db 670 DVLAPEDLTKIKVSTKNGFVTNQISKGQIILKDKIEVLSABDTDDQEKTDSDSD 729

Qy 95 VSKKDDPNVHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNN 142
Db 730 KSKKDDKYDEDNSNASSSKNEKSNADSKNDSDDSTNETSGS--ERNN 775

RESULT 10
US-10-691-672A-7
; Sequence 7, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691.672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 7
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
```

FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)..(189)  
OTHER INFORMATION: MSP3a to MSP3f  
US-10-691-672A-7

Query Match 13.5%; Score 108; DB 5; Length 188;  
Best Local Similarity 23.2%; Pred. No. 0.17;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 17 IQNGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEISGF--EGKKDAG-----YVINLS 69  
DB 15 VLKAKEASS-----YDYL-----GHEFGGVPEHKKEENMLSHLYVSKD 55  
QY 70 KDTFKVPVKKIEKKKE-----ENKPTFDVSKKKDPQVNHSQLNESHKRE 116  
DB 56 KENISKENDVDLDEKEEABETEEBELEKNEETESEISEDEEEEEEKEEKEENDKK 115  
QY 117 DLQREHSQKSDTKDVTATVLDKNNISKSTNN 151  
DB 116 EQEKSQNNNDQKKDMEA-----QNLISKQNNN 145

## RESULT 11

US-10-739-930-6262

Sequence 6262, Application US/10739930  
Publication No. US20040316190A1  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT  
FILE REFERENCE: 38-21(53377) B  
CURRENT APPLICATION NUMBER: US/10/739,930  
CURRENT FILING DATE: 2003-12-18  
NUMBER OF SEQ ID NOS: 11088  
SEQ ID NO 6262  
LENGTH: 470  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
OTHER INFORMATION: Clone ID: ARATH-23APR03-C271270\_1.p  
US-10-739-930-6262

Query Match 13.5%; Score 107.5; DB 5; Length 470;  
Best Local Similarity 20.1%; Pred. No. 0.59;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 9 KPHRVTVIIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEISGFE-----GKK 60  
DB 82 RENRVTDVQNNNGESK-----YVQDLARRIRYDE-EATGCSQAQRIDHPNQK 129  
QY 61 DAGYVINLSKDTFKVPVKKIEKKKEENKPTFDVSKKKDN----- 101  
DB 130 NVGIETKAPENSPIBETSHRVDDNRINQKFTAAKSENAVSFSGADHKRAEVMGK 189  
QY 102 PQVNHSQLNE-----SHRKEDLQREHSQKSDTKDVTATVLDKNNISKSTNNPNK 154  
DB 190 PMENRDQVRQTESAEKSHKENVTYKSEKPRDQGVGKTEAKDKORNEKKEKKEKTESINK 248

## RESULT 12

US-10-691-672A-3

Sequence 3, Application US/10691672A  
Publication No. US20050112133A1  
GENERAL INFORMATION:  
APPLICANT: DRUILHE, PIERRE  
TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND  
TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT  
FILE REFERENCE: 02356.0085  
CURRENT APPLICATION NUMBER: US/10/691,672A  
CURRENT FILING DATE: 2003-10-24  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 3.3

SEQ ID NO 3  
LENGTH: 647  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: Peptide  
NAME/KEY: SITE  
LOCATION: (1)..(647)  
OTHER INFORMATION: GLURP MSP3 fusion protein  
US-10-691-672A-3

Query Match 13.3%; Score 106; DB 5; Length 647;  
Best Local Similarity 22.8%; Pred. No. 1.2;  
Matches 43; Conservative 35; Mismatches 55; Indels 56; Gaps 9;

QY 7 ELKPHRVTVIIONGKEM-----SSTIVSEEDFILPVYKGELEK----- 44  
DB 428 ETVHEBETVQSNEPKADNDGNVSQNNELNENEFV-----ESEKSEHSEARSKAKEA 481  
QY 45 -GYOFD-GWEISGP--EGKKDAG-----YVINLSKDTFKVPVKKIEKKKEENKPTFDV 95  
DB 482 SSYDYLGWEEFGGVPEHKKEENMLSHLYVSKKDKENISKENDVDVLDK-KEEAEETEE 540  
QY 96 SKKKDNQVNHSQLN-----ESHKEDLQREHSQKSDTKDVTATVLDKNN 142  
DB 541 ELEKNEBETETSEISEDEEEEEEKEEKEEKEQNNENNQKKDMEA-----QN 595  
QY 143 ISSKSTNN 151  
DB 596 LISKNQNN 604

## RESULT 13

US-09-820-843A-107  
Sequence 107, Application US/09820843A  
Publication No. US20030039963A1

GENERAL INFORMATION:  
APPLICANT: Council of Scientific and Industrial Research  
TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROT  
TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES  
FILE REFERENCE: Q63915  
CURRENT APPLICATION NUMBER: US/09/820,843A  
CURRENT FILING DATE: 2001-03-30  
NUMBER OF SEQ ID NOS: 118  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 107  
LENGTH: 665  
TYPE: PRT  
ORGANISM: Plasmodium falciparum  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: hypothetical protein  
NAME/KEY: misc feature  
OTHER INFORMATION: gi|3845248  
US-09-820-843A-107

Query Match 13.1%; Score 105; DB 3; Length 665;  
Best Local Similarity 24.4%; Pred. No. 1.5;  
Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;

QY 1 DTGEVSELKPHRVTVIIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI--SGFE 57  
DB 128 EKNKINKSDLHQRQNELNLQSGK-----NQQDI-----NKVEKGKQ-----DISNSAE 170  
QY 58 GKQDAGYVINLSKDTFKVPVKKIEKKKE-----ENKPTFD-----VSKKKDNQK 103  
DB 171 NKQD-----VKEGVKLELEKKEEKKIKSIDDHVKVEENKKSDDHKVEENKKSDDHK 218  
QY 104 VNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSTNNPNK 154  
DB 219 VEENKKSDDHKIEBVKVVEEHEDEE-----DKCKEKKSENKQNDENK 261

```

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52942
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Clostridium difficile
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (18)..(18)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (29)..(29)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (37)..(37)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (43)..(43)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (54)..(54)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (84)..(84)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (96)..(96)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (359)..(359)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (385)..(385)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (388)..(388)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (396)..(396)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (400)..(400)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (402)..(402)
; OTHER INFORMATION: X=any amino acid

```

```
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (404)...(404)
OTHER INFORMATION: X=any amino acid
US-10-282-122A-52942

Query Match
Best Local Similarity 12.9%; Score 103; DB 4; Length 707;
Matches 39; Conservative 25; Mismatches 56; Indels 36; Gaps 5;

QY 1 DTGEVSELPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGKK 60
Db 504 DIGDVVEDKD-----TTDKYDS---NKEDIIPEPNKSKKAKLFG----- 542
QY 61 DAGYVNLKOTFIKPVFKKIEKKEEN--KPTFDVSKKONPQVNHSQLNESHKEDL 118
Db 543 -----FIKKNVEEVEQEEENLNDISFDIILDKPVENNQVKSEIEIQNELKE-I 589
QY 119 QREHHSQSDSKTVDATVTLVDKNNISSKSTTNPNK 154
Db 590 KQEPSQHITEERSVKIEKPIINNLDKVVSSNNESK 625

RESULT 16
US-11-097-143-12723
; Sequence 12723, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12723
; LENGTH: 564
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-12723

Query Match
Best Local Similarity 12.7%; Score 101.5; DB 6; Length 564;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 31 EDFILPVYKGELEKGYQFGM-----RISGPEGKKDAGYVI-----NLSKDTPIK 75
Db 78 EDLDTPLSESRFSK--VFDGWDEHRDHDGHDVQPSGEALDDHDEHDDHDEHDEBE 135
QY 76 PVFKKIEKKEENKPT-----FDVSKKONPQVNHSQLNESHKEDLQREHHSQSDS 129
Db 136 PLTEELEBELEEEPTEDRPADEERYEEDDEENNA--GENITAEADAREEEDNDND 193
QY 130 TKDVTATVTLVDKNNISSKST 148
```

```
Db 194 EGTVEATVEATTEATTEAT 212
RESULT 17
US-10-732-923-18783
; Sequence 18783, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 18783
; LENGTH: 973
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(973)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-18783

Query Match
Best Local Similarity 12.5%; Score 100; DB 5; Length 973;
Matches 32; Conservative 39; Mismatches 69; Indels 6; Gaps 3;

QY 7 ELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGKKDAGYVI 66
Db 597 KLEYTNINMNDQKGECNSTYKT---LIQHRSSKSGKGFHTKINNINREHARGYI 652
QY 67 NLSKDTPIKPVFKKIEKKEENKPTFDVSKKONPQVNHSQLNESHKEDL-QREHHSQ 125
Db 653 SLSKDNNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDN 712
QY 126 KSDSTKQVTVTLVDKNNISSKSTTN 151
Db 713 KSEYTFVTATSNKSKDNDINKSSND 737

RESULT 18
US-10-282-122A-53254
; Sequence 53254, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
```

```

; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53254
; LENGTH: 1184
; TYPE: PRT
; ORGANISM: Clostridium difficile
US-10-282-122A-53254

```

Query Match	12.2%	Score 97.5;	DB 4;	Length 1184;
Best Local Similarity	26.3%	Pred. No. 14;		
Matches	46;	Conservative 30;	Mismatches 70;	Indels 29; Gaps 9;
Qy	4	EVSELEKPHRYVTVTIQ--NGKEMSSITIV--SEEDF--ILPVYKGELEKGVQPDGWEISGPEG	58	
Db	240	ELSEVNEHRKRVIEKELNEKEGQKNVKEKQEDINKEVEVLQDVIEKSVYDIN-SIKGVIS	298	
Qy	59	KQAGYVYNISKQTF-----IKPVFKKIBEKKE-----BENKPTFDVSKKD	100	
Db	299	KCES--QINLIKERIRNFTWEISRKNIKLEIKDIKEKLNENQVYIKESLNK--LSGSEELS	354	
Qy	101	NPQVNHSQLNESHKEDLQREHSQKSDSTKDVDTATVLD--KNNTSSKSTTNNPNK	154	
Db	355	TLOENIKVLEGSKDKQIKLESLNNEITELLESIIIDILNKQOEPSNKLSTLNANK	409	

```

RESULT 19
US-10-732-923-22588
; Sequence 22588, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22588
; LENGTH: 3127
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-22588

```

Query Match	12.1%	Score 97	DB 5	Length 3127
Best Local Similarity	24.0%	Pred. No. 50		
Matches 50	Conservative 32	Mismatches 70	Indels 56	Gaps 8
Qy	1	DTGVSSELKPHRVTVTIQNG-----KEMSSSTIVSEEDFILPVYKGELEKGYQDFDGWEIS	54	
Db	1974	NTSVLSPLHLIGDIVDNNTIKRKKCKKEIKTIIVSDFTSPVNIKEVYNEQERKKEIV	20333	
Qy	55	G---FEGGKAGVYVNLNSKDTFIPVFPVKIKBEKBEENK-----PTF-	93	
Db	2034	GNLSYDGTKKIFPFIKFTKEGRICK--KKIEKKEKKEKNNNNFLYNDYSSYSSPKYG	2091	
Qy	94	-----DVSKGKDNQVNVHSIQI-----NESHK-----EDLQREHSQKS	127	
Db	2092	DNENNFVIKYIRERKDFQKEDPHFNFPKFLHNYNPMKNKCNKCNKCNKCNVRRNEYNYT	2151	

[illegible]

```

RESULT 21
US-10-381-596A-2
; Sequence 2, Application US/10381596A
; Publication No. US20040014178A1
; GENERAL INFORMATION:
; APPLICANT: Biostapro AB
; TITLE OF INVENTION: von Willebrand factor-binding proteins from
; TITLE OF INVENTION: Staphylococci
; FILE REFERENCE: 110059600
; CURRENT APPLICATION NUMBER: US/10/381,596A
; CURRENT FILING DATE: 2003-07-02
; PRIOR APPLICATION NUMBER: SE 0003573--3
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2060
; TYPE: PRT
; ORGANISM: Staphylococcus lugdunensis
US-10-381-596A-2

```

Query Match	12.1%	Score 96.5;	DB 4;	Length 2050;
Best Local Similarity	25.7%;	Pred.No.33;		
Matches	39; Conservative	25; Mismatches	49; Indels	39; Gaps 9;
Qy	4	EVSCLKPHRVTVTIQ--NGKMSSTIVSEEDPILPVYGELEKGVPFG--WEISGPEKK	60	
Db	1914	DIDELK--ITIIYDVTNGRE-----IVPSRKQLPP-EQFIGQWQYTGHK---	1955	
Qy	61	DAGVIVNLSDTFIKPVFKIEB-----KKEENKPTFDVSKCKNDPNVHSQLNESH	113	
Db	1956	-----IEKDGIITYIYKKEVNAPAKLQKTKHN--TQSEGFRETPOVKQLVKYNH	2006	





```
; OTHER INFORMATION: Clone ID: PAT_MRT4530_25224C.1.pcp
US-10-437-963-122282

Query Match      11.6%; Score 93; DB 4; Length 869;
Best Local Similarity 45.1%; Pred. No. 23;
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;

QY 94 DVSKKKDNPVNHSQLSHRKEQLQREHSQKSDTKDVTATVLDKNN 142
Db 617 DASKRDNHQSEGNL--SHRDEPTRKKKQKTKNATSDACAQEVVTEKNN 665

RESULT 26
US-10-732-923-8762
; Sequence 8762, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8762
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1529)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-8762

Query Match      11.6%; Score 93; DB 5; Length 1529;
Best Local Similarity 27.2%; Pred. No. 47;
Matches 28; Conservative 19; Mismatches 30; Indels 26; Gaps 5;

QY 67 NLSKDT-----FIKPVFKKIEKKEENKPTFDVSKKDNPOVNHSQL-----NES 112
Db 356 NIISDTQCIKIPKINSEYKKEKKEBK-----NEKINDTIHYSESISKNSDNEQ 409

QY 113 H-----RKEDLQREHSQKSDTKDVTATVLDKNN--ISSKSTT 149
Db 410 HPFLSKLRNVKBEKKKKKKTKTKTKVIAQKNKTVIAQKNKT 452

RESULT 27
US-10-282-122A-52328
; Sequence 52328, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELTRA 03A4
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

; OTHER INFORMATION: Clone ID: PAT_MRT4530_25224C.1.pcp
US-10-437-963-122282

Query Match      11.6%; Score 93; DB 4; Length 869;
Best Local Similarity 45.1%; Pred. No. 23;
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;

QY 94 DVSKKKDNPVNHSQLSHRKEQLQREHSQKSDTKDVTATVLDKNN 142
Db 617 DASKRDNHQSEGNL--SHRDEPTRKKKQKTKNATSDACAQEVVTEKNN 665

RESULT 26
US-10-732-923-8762
; Sequence 8762, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8762
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1529)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-8762

Query Match      11.6%; Score 92.5; DB 4; Length 903;
Best Local Similarity 25.5%; Pred. No. 27;
Matches 42; Conservative 27; Mismatches 57; Indels 39; Gaps 9;

QY 2 TGEVSELKPHR---VTVTIQNGKEMSTIVSEBDFILPVYKGELEKGYQPDGWEISQFE 57
Db 173 SNESIELKNEKSKQLPKINVELKEEKSKQIKKEHNL-----KNEKSKSIKPKSIXL---K 225
QY 58 GKGDAGYVINLSKDTFIKPVFKKIEKKEENKPTFDVS-----KKKDNPOVNHSQLNE 111
Db 226 EEKD-----KQVPFKEHNSENEENKQISKVNVKLNERSKQLPKIN-VELKE 271
QY 112 SHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN--PNK 154
Db 272 EKNQSI--KEHNELREBETK----KCLPKVNIELKEBETKKQVFNK 310

RESULT 28
US-10-755-889-615
; Sequence 615, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 615
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-615

Query Match      11.6%; Score 92.5; DB 4; Length 2468;
Best Local Similarity 31.0%; Pred. No. 93;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 68 LSKDTFIKPVFKKIEKKEENKPTFDVSKKDNPOVNHSQLSHRKEQLQRE-----E 122
Db 638 VKETKVKP-----EDKKEKEKPKKEVAKKEDKTFI---KKEBKPKKEEVKKEVKBK 689
```

QY 123 HSQKSDSTKDV 133  
Db 690 KEEKKEPKKEV 700

## RESULT 29

US-10-489-740-216  
; Sequence 216, Application US/10489740  
; Publication No. US20050112574A1  
; GENERAL INFORMATION:  
; APPLICANT: Bionomics Limited  
; TITLE OF INVENTION: P9  
; FILE REFERENCE: Angiogenesis PCT  
; CURRENT APPLICATION NUMBER: US/10/489,740  
; CURRENT FILING DATE: 2004-03-15  
; NUMBER OF SEQ ID NOS: 216  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 216  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-489-740-216

Query Match 11.6%; Score 92.5; DB 5; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 93;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 68 LSKDTFIKPVFKKEEKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQRE-----E 122  
Db 638 VKKETKVP-----EDKKEEKEPKKEVAKKEDKTPI---KKEEKPKEEKKVKKKEIK 689

QY 123 HSQKSDSTKDV 133  
Db 690 KEEKKEPKKEV 700

## RESULT 30

US-10-450-763-46995  
; Sequence 46995, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc

; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

; FILE REFERENCE: 790CIEP/US

; CURRENT APPLICATION NUMBER: US/10/450,763

; CURRENT FILING DATE: 2003-06-11

; PRIOR APPLICATION NUMBER: PCT/US01/08631

; PRIOR FILING DATE: 2001-03-30

; PRIOR APPLICATION NUMBER: 09/540,217

; PRIOR FILING DATE: 2000-03-31

; PRIOR APPLICATION NUMBER: 09/649,167

; PRIOR FILING DATE: 2000-08-23

; NUMBER OF SEQ ID NOS: 60736

; SOFTWARE: Custom

; SEQ ID NO 46995

; LENGTH: 2519

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: DOMAIN

; LOCATION: (1040)..(1091)

; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,

; OTHER INFORMATION: accession number BL00412D, p-value=2.432e-11, raw score of 16.54

; FEATURE:

; NAME/KEY: DOMAIN

; LOCATION: (1919)..(2122)

; OTHER INFORMATION: (1919)..(2122)

; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by PFam,

; OTHER INFORMATION: accession name MAP1B\_neuraxin, E-values=1.9e-59, PFam score of 190

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)...(2519)

; OTHER INFORMATION: Xaa = X or \* as defined in Table 2

US-10-450-763-46995

Query Match 11.6%; Score 92.5; DB 5; Length 2519;  
Best Local Similarity 31.0%; Pred. No. 95;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 68 LSKDTFIKPVFKKEEKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQRE-----E 122  
Db 689 VKKETKVP-----EDKKEEKEPKKEVAKKEDKTPI---KKEEKPKEEKKVKKKEIK 740

QY 123 HSQKSDSTKDV 133  
Db 741 KEEKKEPKKEV 751

## RESULT 31

US-10-437-963-187665  
; Sequence 187665, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 187665

; LENGTH: 1005

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_84346C.1.pep

US-10-437-963-187665

Query Match 11.5%; Score 92; DB 4; Length 1005;  
Best Local Similarity 26.4%; Pred. No. 34;  
Matches 32; Conservative 22; Mismatches 41; Indels 26; Gaps 4;

QY 55 GFEGKDGAGVYVNLKDTFIKPVFKKEEKEENKPTFDVSKK-KNPQV--NHSOLNE 111

Db 35 GSEKKEMSGKNKSIKETGTGQSKELQKESKRSKSTKDKSKKNKDMTQVPTNABPHK 94

QY 112 SH-----RKE-----DLQREHSQKSDSTKDVTTATVLDKNNISSKST 148

Db 95 EYTTKVKRESRTDSSIEQVIGTSSIQEMETNEQVSKSDTSKDMTQVPANAEGIRKEYT 154

QY 149 T 149  
Db 155 T 155

## RESULT 32

US-10-732-923-22709  
; Sequence 22709, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:

; APPLICANT: Edgerton, Michael D

; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES

; FILE REFERENCE: 38-15(52796)C

; CURRENT APPLICATION NUMBER: US/10/732,923

; CURRENT FILING DATE: 2003-12-10

; PRIOR APPLICATION NUMBER: 10/310,154

; PRIOR FILING DATE: 2002-12-04

; NUMBER OF SEQ ID NOS: 24149

; SEQ ID NO 22709

; LENGTH: 3124

```
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-22709

Query Match      11.5%; Score 91.5; DB 5; Length 3124;
Best Local Similarity 23.4%; Pred. No. 1.5e+02;
Matches 48; Conservative 31; Mismatches 73; Indels 53; Gaps 7;

QY 1 DTGEVSELKPHRVTVTIQNG-----KEMSSIVSEEDFILPVYKGELEKGYQFDGWEIS 54
Db 1974 NTSYVLESPLHLIGDIVDNINIKRKKKKKEIKTIVSDMFTSPVNIKEYNYEQERKGIIV 2033
QY 55 G---PEGKKDAGVINLSKDTFIKVPFKKIEKKEENKPTF----- 93
Db 2034 GNLSDYTKYKICFPFKFKYKEGRKK--NKIEKKEKYNPNFLYNDYSSSPKYGNE 2091
QY 94 -----DVSKKKDNPOVNHSQL-----NESHKKE-----DLQREHSQKSDST 130
Db 2092 NNFVIKIRERKDFQCKFDHPNPNFSKFLHNYNPMKNKNKNKNVNVVRNEYPNYSSTSS 2151
QY 131 KD-VTATVLDKNISKSSTNNPNK 154
Db 2152 KDGVSYNFLSDSLFSSDNEYSDDNE 2176

RESULT 33
US-10-171-311-188
; Sequence 188, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerh, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 188
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-188

Query Match      11.3%; Score 90.5; DB 4; Length 743;
Best Local Similarity 23.8%; Pred. No. 32;
Matches 38; Conservative 34; Mismatches 51; Indels 37; Gaps 7;

QY 7 ELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGE-----LEKGYQFDGWEI--SGF 56
Db 281 EARPRR-----QSMKEKEHQVVRNEEHKABQEGKVAQREELVETGNHNDVEIEEAGE 335
QY 57 EGKDKDAGVINLSKDTFIKVPFKKIEKKEENKPTFQVSKKDNPNQVNHQVNHSHKRE 116
Db 336 EKEKIGIVHSDAE-----KEQEEKQKQMEVMEKEE-----TEVRESEKQQ 378
QY 117 DLQREHS---QKSDSTKDVTA--TVLDKNISKSSTNN 151
Db 379 DSQPBEVMDVLMVENVKHVIADQEVMTNRVESVEPSN 418
```

```
RESULT 34
US-09-925-299-859
; Sequence 859, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 859
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (590)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-859

Query Match      11.3%; Score 90.5; DB 3; Length 758;
Best Local Similarity 28.1%; Pred. No. 33;
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 3 GEVSELKPHRVTVTIQNGKEMSSIVSE--EDFILPVYKGELEKGYOF-----DGWE--- 52
Db 399 GELQQLSGSQL-----HGKSDSPNVYTEKKKEIATLRRLTELEKRLTFEQQRSDLWERLY 453
QY 53 -----ISGFEGKKDAG---YVINLSKDTFIKVPFKKIEKKEENKPTFQVSKKDD 100
Db 454 VEAKDQNGKQTDGKKKGGRSHRAKNKSKETPLGSV-----KETFDAMKNST 501
QY 101 NPQVNH-----SQLNESHKREKEDLQREHSQKSD--STKQVATVLDK 140
Db 502 KSFVRHHEKIKQAKEA--VKENLKKFSDSVKSTPRHFQKDTTKNIFDE 547

RESULT 35
US-09-925-299-859
; Sequence 859, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 859
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (590)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-859

Query Match      11.3%; Score 90.5; DB 3; Length 758;
Best Local Similarity 28.1%; Pred. No. 33;
```

Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;  
QY 3 GEVSELKPHRVTVTQNGKEMSTVSR--EDFILPVYKLEKGYQF-----DQWE--- 52  
DB 399 GELQQLSGSQL-----HGKSDSPNVYTEKKEITAILERLTELKLTLPQQRSDLWERLY 453  
QY 53 -----1SGPEGKQDAG-----YVNLKSDTFIKPVFKKIEBKKEENKPTFDVSKKD 100  
DB 454 VRANDQKQGTGDKKGGKGRSHRAKNKSKTFLGSV-----KETFDAMKNT 501  
QY 101 NPQVNH-----SQLNESHKEDLQREHSQKSD--STKDVATATVLDK 140  
DB 502 KEFVRHHEKIKQAKA-KVENLAKFSDSVKSTFRHFKDTTKNIIDE 547

## RESULT 36

US-10-282-122A-52510  
; Sequence 52510, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 52510

; LENGTH: 932

; TYPE: PRT

; ORGANISM: Clostridium botulinum

US-10-282-122A-52510

Query Match 11.3%; Score 90.5; DB 4; Length 932;  
Best Local Similarity 24.2%; Pred. No. 42;  
Matches 40; Conservative 30; Mismatches 48; Indels 47; Gaps 9;

QY 5 VSELKPHRVTVTQNGKEMSTVSR--EDFILPVYKLEKGYQF--DQWEISGPEGKQDAGY 64  
DB 742 IEGVKNYSVS-----ELENEIV-----FLRKTIKGGADQSY-----GIEVAKLAGL 784

QY 65 ---VINLSKDTFIKPVFKKIEBKKEE-----ENKPTFDVSKKKDN-----POV 104  
DB 785 PSPVINRAKE-----ILQHIQEDKEENSLNAPSKEYKSDYIEVSKDTSNTKNLGSSEI 839  
QY 105 NLSOLNESHK---EDLQREHSQKSDSTKDVATATVLDKNNISSK 146  
DB 840 KHDYLSNTDTTIIEDSTKEH--LSSNKQINCRINDEKSIKKE 882

## RESULT 37

US-10-724-972A-5858  
; Sequence 5858, Application US/10724972A  
; Publication No. US20040147734A1  
; GENERAL INFORMATION:

; APPLICANT: Doucette-Stamm, Lynn

; APPLICANT: Bush, David

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: PATH03-16

; CURRENT APPLICATION NUMBER: US/10/724,972A

; CURRENT FILING DATE: 2003-12-01

; PRIOR APPLICATION NUMBER: 09/450,969

; PRIOR FILING DATE: 1999-11-29

; PRIOR APPLICATION NUMBER: 09/134,001

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 7544

; SEQ ID NO 5858

; LENGTH: 442

; TYPE: PRT

; ORGANISM: S.epidermidis

US-10-724-972A-5858

Query Match 11.3%; Score 90; DB 4; Length 442;  
Best Local Similarity 21.8%; Pred. No. 19;  
Matches 49; Conservative 29; Mismatches 65; Indels 82; Gaps 11;

QY 1 DTGEVSELKPHRVTVT--IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 44  
DB 91 DT-EINPVAQEVNVTQIEKGKDFEATVTVPEVKLGDKYGLGIEKQETDLSDEELQE 149

QY 45 -----GVQFDGWEISG--PEGKQDAGVINLSKDTFIKP 76  
DB 150 SIDHLSHLAEMVVKEDGAVNGTVDNIDFSG-SVDSEEPDGGQAGYDLEIGSGSFIG 208

QY 77 VFKKIEBKKEENK-----PTFDVS-----KKONPQVNHSQLNE-- 111  
DB 209 FEEQIEGKMGTDGKDVVVTTPPEYHABELAGKEATFKTVNEIKFKDVPFLNDEIANELD 268

QY 112 -----SHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTN 150  
DB 269 SDAENVDEYKENLKRKLSQKATEAENT-----EKBEAINKATEN 308

## RESULT 38

US-10-282-122A-70294  
; Sequence 70294, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 23.9441 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773  
Perfect score: 799  
Sequence: 1 DTGEVSELKPHRVTTIQNG.....ATVLDKNNISSKSTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5 COMB.pdp.\*  
2: /cgn2\_6/ptodata/1/1aa/6 COMB.pdp.\*  
3: /cgn2\_6/ptodata/1/1aa/H COMB.pdp.\*  
4: /cgn2\_6/ptodata/1/1aa/PCUTUS COMB.pdp.\*  
5: /cgn2\_6/ptodata/1/1aa/RE COMB.pdp.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	799	100.0	773	2	US-09-590-991-8
2	796	99.6	637	2	US-09-107-433-3169
3	796	99.6	2138	2	US-09-583-110-5274
4	615	77.0	117	2	US-08-961-083-68
5	615	77.0	117	2	US-09-536-784-68
6	615	77.0	117	2	US-09-765-271-68
7	615	77.0	117	2	US-09-765-272A-68
8	118	14.8	746	2	US-09-710-279-652
9	118	14.8	778	2	US-09-134-001C-3868
10	101.5	12.7	347	2	US-09-248-796A-16224
11	95	11.9	348	2	US-09-538-092-1316
12	92.5	11.6	2468	2	US-09-976-594-726
13	92.5	11.6	2468	2	US-09-538-092-1135
14	92.5	11.6	2522	2	US-09-949-016-10237
15	90.5	11.3	743	2	US-08-910-925-3
16	90.5	11.3	743	2	US-09-949-016-6261
17	90.5	11.3	758	2	US-09-949-016-8288
18	90	11.3	442	2	US-09-134-001C-3033
19	90	11.3	654	2	US-10-172-502-10
20	89.5	11.2	402	2	US-09-464-483-4
21	89.5	11.2	402	2	US-09-414-664-4
22	89.5	11.2	511	2	US-09-198-452A-509
23	89.5	11.2	511	2	US-09-438-185A-475
24	89.5	11.2	529	2	US-09-464-483-2
25	89.5	11.2	529	2	US-09-414-664-2
26	89	11.1	280	2	US-09-248-796A-17646
27	87.5	11.0	1702	2	US-08-296-791-5

28	87.5	11.0	1702	2	US-09-839-996-5	Sequence 5, Appli
29	87.5	11.0	1702	2	US-10-080-505-5	Sequence 5, Appli
30	87.5	11.0	1702	2	US-10-645-655-5	Sequence 5, Appli
31	87.5	11.0	1702	4	PCT-US95-10661A-5	Sequence 5, Appli
32	87	10.9	902	2	US-09-134-001C-5157	Sequence 5157, Ap
33	86	10.8	172	2	US-09-248-796A-21065	Sequence 21065, A
34	86	10.8	243	2	US-09-248-796A-20306	Sequence 20306, A
35	86	10.8	262	2	US-09-248-796A-21451	Sequence 21451, A
36	86	10.8	700	2	US-08-235-836C-74	Sequence 74, Appli
37	86	10.8	1315	2	US-09-200-650B-5	Sequence 5, Appli
38	86	10.8	1989	2	US-09-949-016-10076	Sequence 10076, A
39	85.5	10.7	109	2	US-09-248-796A-24668	Sequence 24668, A
40	85.5	10.7	1007	2	US-09-538-092-736	Sequence 736, App
41	85	10.6	465	2	US-09-134-001C-3856	Sequence 3856, Ap
42	85	10.6	472	2	US-09-710-279-658	Sequence 658, App
43	85	10.6	700	1	US-07-720-589-2	Sequence 2, Appli
44	85	10.6	700	1	US-08-785-190-2	Sequence 2, Appli
45	85	10.6	700	2	US-08-235-836C-66	Sequence 66, Appli

ALIGNMENTS

RESULT 1  
US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%;	Score	799;	DB	2;	Length	773;
Best Local Similarity	100.0%;	Pred. No.	3.8e-75;				
Matches	154;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	DTGEVSELKPHRVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK	60				
Db	620	DTGEVSELKPHRVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK	679				
Qy	61	DAGVINLSKDTFTKPVFKLEEKEENKPTFDVSKKNDPNQVNHSLNESHRKEDLQR	120				
Db	680	DAGVINLSKDTFTKPVFKLEEKEENKPTFDVSKKNDPNQVNHSLNESHRKEDLQR	739				
Qy	121	EEHSQKSDTKDVTATVLDKNNISSKSTNNPNK	154				
Db	740	EEHSQKSDTKDVTATVLDKNNISSKSTNNPNK	773				

RESULT 2  
US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-433-3169

Query Match          99.6%; Score 796; DB 2; Length 637;
Best Local Similarity 99.4%; Pred. No. 6.1e-75;
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60
DB 450 DTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 509
QY 61 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHSQLNESHKEDLQR 120
DB 510 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHSQLNESHKEDLQR 569
QY 121 EBHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 154
DB 570 EDHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 603

RESULT 3
US-09-583-110-5274
; Sequence 5274, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
```

```
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5274
; LENGTH: 2138
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5274

Query Match          99.6%; Score 796; DB 2; Length 2138;
Best Local Similarity 99.4%; Pred. No. 3.2e-74;
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60
DB 1951 DTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 2010
QY 61 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHSQLNESHKEDLQR 120
DB 2011 DAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKDNPNVHSQLNESHKEDLQR 2070
QY 121 EBHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 154
DB 2071 EDHSQKSDSTKDVTTATVLDKNNISSKSTNNPNK 2104

RESULT 4
US-08-961-083-68
; Sequence 68, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-68

Query Match          77.0%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.9e-57;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 YKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 97
```



Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDTFKVPFKIEEKEENKPTFDVSK 60  
QY 98 KKDNPQVNHSQLNESHKEDLQREHRSQKSDTKDVTATVLDKNNISSKSTNNPNK 154  
Db 61 KKDNPQVNHSQLNESHKEDLQREHRSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

## RESULT 5

US-09-536-784-68  
; Sequence 68, Application US/09536784  
; Patent No. 6573082

## GENERAL INFORMATION:

APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/536,784  
FILING DATE: 30-Oct-1997  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083  
FILING DATE: OCT-30-1997

## ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

## REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB340P3

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

## INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:

LENGTH: 117 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 68:

US-09-536-784-68

Query Match 77.0%; Score 615; DB 2; Length 117;

Best Local Similarity 100.0%; Pred. No. 5.9e-57;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDTFKVPFKIEEKEENKPTFDVSK 97

Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDTFKVPFKIEEKEENKPTFDVSK 60

QY 98 KKDNPQVNHSQLNESHKEDLQREHRSQKSDTKDVTATVLDKNNISSKSTNNPNK 154

Db 61 KKDNPQVNHSQLNESHKEDLQREHRSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

## RESULT 6

US-09-765-271-68  
; Sequence 68, Application US/09765271  
; Patent No. 6887663

## GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,271  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/536,784

FILING DATE: <Unknown>

APPLICATION NUMBER: 08/961,083

FILING DATE: OCT-30-1997

## ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB340P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

## INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:

LENGTH: 117 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 68:

US-09-765-271-68

Query Match 77.0%; Score 615; DB 2; Length 117;

Best Local Similarity 100.0%; Pred. No. 5.9e-57;

Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDTFKVPFKIEEKEENKPTFDVSK 97

Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDTFKVPFKIEEKEENKPTFDVSK 60

QY 98 KKDNPQVNHSQLNESHKEDLQREHRSQKSDTKDVTATVLDKNNISSKSTNNPNK 154

Db 61 KKDNPQVNHSQLNESHKEDLQREHRSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

## RESULT 7

US-09-765-272A-68

; Sequence 68, Application US/09765272A

; Patent No. 6929930

## GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and

NUMBER OF SEQUENCES: 454

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: Dell Latitude C610

OPERATING SYSTEM: Windows 2000

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

```

; SEQ ID NO 16224
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16224

Query Match      12.7%   Score 101.5; DB 2; Length 347;
Best Local Similarity 21.4%; Pred. No. 0.02;
Matches 31; Conservative 32; Mismatches 63; Indels 19; Gaps 0

QY    22 EMSSTIVSEEDFILPVYKGELEKGQFDGWEISGF-----EGKKDAGVYINLSKDT
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     202 KLSPLMDEIDTTPFYND--KKWYVATSSLQNVYQTDLESSESGCWEDDLLENKN
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY    76 PVFKIKIEKEEENKPTFVSKCKCNQPQNHSQLN-----ESHKEDLQREHSHSQK

```

Db 260 PVFKTLDLQREE-----WKAEREQANPKKEEENLNQKPVAKQKQPNSTKQKQKQKQ 314  
QY 130 TKDVTATVLDKNNISKSSTNNPNK 154  
Db 315 TKKITPKTSKRWLEGISISNLIK 339

RESULT 11  
US-09-538-092-1316  
; Sequence 1316, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 1316  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number Q14093  
US-09-538-092-1316

Query Match 11.9%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 0.096;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;  
QY 39 KGELEKGYQFGWEISGPGKKDAGVINLSKDTFKPVFKKIEEKGEEENKPTF---DV 95  
Db 205 ESEGEKG-----GTEKDSKKGKDS-----KXGKDSAIELQAVKADSKEDGKDKANGDE 256  
QY 96 SK-KKQNPVNHSQLN-----ESHRKEDLQREHSQKSDSTKD---VTATVLDKNNI 143  
Db 257 SKDAKDAKEIKKGKDKKKPSSTDSDSKDVKE---SKDATKDAKKVAKKDTKEGSA 313  
QY 144 SSK 146  
Db 314 DSK 316

RESULT 12  
US-09-976-594-726  
; Sequence 726, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 726  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature

; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1  
US-09-976-594-726  
Query Match 11.6%; Score 92.5; DB 2; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 2.5;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
QY 68 LSKDTFKPVFKKIEEKGEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQRE-----E 122  
Db 638 VKKETKVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKEEKPKEEKKVKKGRK 689  
QY 123 HSQKSDSTKDV 133  
Db 690 KEEKKEPKKEV 700

RESULT 13  
US-09-538-092-1135  
; Sequence 1135, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 1135  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P46821  
US-09-538-092-1135

Query Match 11.6%; Score 92.5; DB 2; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 2.5;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
QY 68 LSKDTFKPVFKKIEEKGEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQRE-----E 122  
Db 638 VKKETKVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKEEKPKEEKKVKKGRK 689  
QY 123 HSQKSDSTKDV 133  
Db 690 KEEKKEPKKEV 700

RESULT 14  
US-09-949-016-10237  
; Sequence 10237, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08



```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8288
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8288

Query Match      11.3%; Score 90.5; DB 2; Length 758;
Best Local Similarity 28.1%; Pred. No. 0.82;
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 3 GEVSELKPHRVVTVTQNGKEMSTIVSE--EDFILPVYKGELEKGYQP-----DWE--- 52
Db 399 GELQQLSGSL-----HGKSDSPNVVTEKKEKAILRLTELRLTELEKLTPEQQRSDLWERLY 453
QY 53 -----ISGPEGKKDAG-----YVINLSKDTFIKVPKKBIEKKEEENKPTFDVSKKQD 100
Db 454 VEAKQNGKQGTGDKKKGGRSHRAKNKSKETFLGSV-----KETFDANKST 501
QY 101 NPQVNH-----SOLNESHKEDLQREHSQKSD--STKDVTVATVLDK 140
Db 502 KEFVRHHEKIKQAKEA--VKENLKKFSDSVKSTFRHFKDTTKNIFDE 547

RESULT 18
US-09-134-001C-3033
; Sequence 3033, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3033
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3033

Query Match      11.3%; Score 90; DB 2; Length 442;
Best Local Similarity 21.8%; Pred. No. 0.45;
Matches 49; Conservative 29; Mismatches 65; Indels 82; Gaps 11;

QY 1 DTGVSSELKPHRVTVT--IQNGKE--MSSTIVSEEDFILPVYKG--BLEK----- 44
Db 91 DT--EINPVAQPEVNTQIEKGDFEATVTVPEPVKLDYGLIEKQETDLSDEELQE 149
QY 45 -----GYQFDGWEISG--FEKKDAGVYVNLKDTPTPK 76
Db 150 SIDHSLSLHAEVVKEDGAVNGDVTVNDIFSG--SVDGSEFDGQAEYDLGSGSFIPG 208
QY 77 VPKKIEKKEEENK-----PTFDVS-----KKKDNPOVNHSQLNE-- 111
Db 209 FEEQIEGKMTGDEKDVVVTFPEYVHAEELAGKEATFKTVKNIEKPKDVPFELNDEIANE 268
QY 112 -----SHRKDLQREHSQKSDTKDVTATVLDKNNISSKSTTN 150
Db 269 SDAENVDEYKLNRLKRLSEKQATEAENT-----EKEEAINKATEN 308

RESULT 19
US-10-172-502-10
; Sequence 10, Application US/10172502
; Patent No. 6841154
; GENERAL INFORMATION:
; APPLICANT: POSTER, Timothy et al
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES...
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-10

Query Match      11.3%; Score 90; DB 2; Length 654;
Best Local Similarity 21.4%; Pred. No. 0.76;
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;

QY 4 EVSELKPHRVTV--TIQNGKEMSTIVSEEDFILPVYKGELE--KGYQP-----DGWEIS 54
Db 350 KMTDLQDTKYVYVY2SVENNESMDTFFVKH-----PIKTGMLNGKKYVMVMTTNDYWKDF 404
QY 55 GPEGKK-----DAGVYVNL--SKDTFI 74
Db 405 MVEGQVRVTISKDAKNNTRTIIPVYVEGKTYDAIVKVHVVTIDYDQGHVRIVDKEAPT 464
QY 75 KPVFKKIBEEKKEENKPTFDV-----SKKDNPOVNHSQLNESHKEDLQ-----REHSQ 125
Db 465 KANTDKSNKKQODNSAKKEATPATPSKPTSPVEKESQKQDSQKQDNDKQLPSVEKENDA 524
QY 136 KSDSTQVT--ATVLDKNNISSKSTTNPNK 154
Db 525 SSESQKDTPATKPTKGEVSSSTT--PTK 552

RESULT 20
US-09-464-483-4
; Sequence 4, Application US/09464483
; Patent No. 6228617
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6228617el tig
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/464,483
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,339
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-464-483-4

Query Match      11.2%; Score 89.5; DB 2; Length 402;
Best Local Similarity 22.1%; Pred. No. 0.44;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 1 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 44
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 DETDIKPAQPEVSVTQIEKGKDFIPEATVTVPEVKLGDYKGLIEIKQETELSDDELQ 112
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 45 -----GYQPDGWEISG--FEGKKDAGVIVNLSKDTTFIKP 76
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 AIDHSLGHLAEMVVKEDGVWNGDTWNIDFSG-SVDGEEFEGQAEGYDLEIGSGSPI-P 170
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 77 VFK-----KIEEKK-----EE--NKPTFDVS-----KKKNPQVNHSQLNE- 111
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 GFEEQLEGKMKVDEKDVVVTPPEYHAEELAGKEATFKTKVNEIKFKEVPELTDEIANEL 230
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 112 -----SHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 150
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 DAEANTVDEYKENLRKRLAEQKATDAENV-----EKEEAITKATDN 271
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 22
US-09-198-452A-509
; Sequence 509, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 509
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-509

Query Match      11.2%; Score 89.5; DB 2; Length 511;
Best Local Similarity 24.5%; Pred. No. 0.61;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 74 IKPVFKKIEKKKEENKPTFD-----VSKKDNPNQVNHSQLNESHRK 115
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 VKGVFKKTPQARPEVSSPRLPSHVHQRLPGLGFRDRIQKRSENPEADLGKMKRSYSD 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 149
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 GDLDRVGHDSNEDSTEDSRS---EGGEFPSSKSS 185
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 23
US-09-438-185A-475
; Sequence 475, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
; APPLICANT: Stephens, Richard
; APPLICANT: Mitchell, Wayne
; APPLICANT: Kalman, Sue
; APPLICANT: Davis, Ronald
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
; FILE REFERENCE: 018941-000411US
; CURRENT APPLICATION NUMBER: US/09/438,185A
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: US 60/108,279
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: US 60/128,606
; PRIOR FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 1074
; SOFTWARE: FastSeq for Windows Versicn 3.0
; SEQ ID NO 475
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
```



```
US-09-248-796A-17646
; Sequence 17646, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17646
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17646

Query Match      11.1%; Score 89; DB 2; Length 280;
Best Local Similarity 25.9%; Pred. No. 0.3;
Matches 42; Conservative 25; Mismatches 55; Indels 40; Gaps 10;

QY 3 GEVSELKPHRVTVTIQNGKEMSTIVSEDFILPVYKGELEKGYQFDGWEISFGEGKDA 62
DB 72 GHVSS-----TYTVKRT--SVASTFCSKYDFNVFSYANLSLGF-----ELYSYANKK-- 116

QY 63 GYVINLSKDTFKIPVKFKIE-EKKEEENKPTFDVSKKKONPQVNHVS-QLNESHKREDLQR 120
DB 117 -----KNSF--PSEFHEIHISSSEENK-----YLKGPPELQRLHNLHLHQVRPIKS 163

QY 121 EHS-----QKSDSTKDVATVLDKNNISSKSTNNPN 153
DB 164 HKYEGNRTIINPQNLNDVYHINPTLLSSNG-STSTNNEN 204

RESULT 27
US-08-296-791-5
; Sequence 5, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; APPLICATION NUMBER: US/08/296,791
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-08-296-791-5

Query Match      11.0%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 5.1;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 66 INLSKDTFKIPVKFKIEEKEENKPTFDVSKKKONPQVNHVSQLNESHKREDLQREHS- 124
DB 1296 INTGSATAITETAESDKPQTETAASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY 125 --QKSDSTKDVATVLDKNNISSKSTNNPNK 154
DB 1356 SQPQTSABETTAATDDETTIADNSGKSKENR 1387

US-09-839-996-5
; Sequence 5, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-839-996-5

Query Match      11.0%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 5.1;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 66 INLSKDTFKIPVKFKIEEKEENKPTFDVSKKKONPQVNHVSQLNESHKREDLQREHS- 124
DB 1296 INTGSATAITETAESDKPQTETAASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY 125 --QKSDSTKDVATVLDKNNISSKSTNNPNK 154
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
US-08-296-791-5

Query Match      11.0%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 5.1;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 66 INLSKDTFKIPVKFKIEEKEENKPTFDVSKKKONPQVNHVSQLNESHKREDLQREHS- 124
DB 1296 INTGSATAITETAESDKPQTETAASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY 125 --QKSDSTKDVATVLDKNNISSKSTNNPNK 154
DB 1356 SQPQTSABETTAATDDETTIADNSGKSKENR 1387

RESULT 28
US-09-839-996-5
; Sequence 5, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-839-996-5

Query Match      11.0%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 5.1;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 66 INLSKDTFKIPVKFKIEEKEENKPTFDVSKKKONPQVNHVSQLNESHKREDLQREHS- 124
DB 1296 INTGSATAITETAESDKPQTETAASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY 125 --QKSDSTKDVATVLDKNNISSKSTNNPNK 154
```



[illegible]

Qy	125	--QKSDTKDVTATVLDKNNISSKSTNNPNK	154
Db	1356	SQOETSAEETAASTDDTTIADNSKSKPNR	1387
RESULT 32			
US-09-134-001C-5157			
; Sequence 5157, Application US/09134001C			
; Patent No. 6380370			
; GENERAL INFORMATION:			
; APPLICANT: Lyun Doucette-Stamm et al			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS			
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: GTC-007			
; CURRENT APPLICATION NUMBER: US/09/134.001C			
; CURRENT FILING DATE: 1998-08-13			
; PRIOR APPLICATION NUMBER: US 60/064,964			
; PRIOR FILING DATE: 1997-11-08			
; PRIOR APPLICATION NUMBER: US 60/055,779			
; PRIOR FILING DATE: 1997-08-14			
; NUMBER OF SEQ ID NOS: 5674			
; SEQ ID NO 5157			
; LENGTH: 902			
; TYPE: PRT			
; ORGANISM: Staphylococcus epidermidis			
US-09-134-001C-5157			
Query Match 10.9%; Score 87; DB 2; Length 902;			
Best Local Similarity 25.3%; Pred. No. 2.4;			
Matches 38; Conservative 27; Mismatches 65; Indels 20; Gaps 7;			
Qy	14	TVTTONKEMS-STIVSEEDFILPVYKELEKGYQFDGWIEISGPEGKKDAGY-VINLSKD	71
Db	755	TITERNGNIVCITTVTGEEDILMVVTNAGVI---IRLDVHDISQ-NGRAAQGVRLMKLGDG	810
Qy	72	TFIKPVFKTTEKKEEN-----KPTFDVSKKDNPOVNHSQLNESHKEDLQREE	122
Db	811	QFVSTVAKVNEEDNEENADAAQOSTTTETADVSEVD----DTPPGNAIHTEGDAEMES	866
Qy	123	-HSQKSDTKDVTATVLDKNNISSKSTNN	151
Db	867	VESPENDDRIDRQDFMDRVNEDIESADN	896
RESULT 33			
US-09-248-796A-21065			
; Sequence 21065, Application US/09248796A			
; Patent No. 6747137			
; GENERAL INFORMATION:			
; APPLICANT: Keith Weinstock et al			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS			
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: 107196.132			
; CURRENT APPLICATION NUMBER: US/09/248,796A			
; CURRENT FILING DATE: 1999-02-12			
; PRIOR APPLICATION NUMBER: US 60/074,725			
; PRIOR FILING DATE: 1998-02-13			
; PRIOR APPLICATION NUMBER: US 60/096,409			
; PRIOR FILING DATE: 1998-08-13			
; NUMBER OF SEQ ID NOS: 28208			
; SEQ ID NO 21065			
; LENGTH: 1702			
; TYPE: PRT			
; ORGANISM: Candida albicans			
US-09-248-796A-21065			
Query Match 10.8%; Score 86; DB 2; Length 172;			
Best Local Similarity 22.1%; Pred. No. 0.32;			
Matches 34; Conservative 25; Mismatches 67; Indels 28; Gaps 4;			
Qy	11	HRVTVT--IQNGKMSSTIVSEED-----FILPVYKGL-----EKGYQFDGWIEISGF	56

QY 47 QFDGWBIS

QY 47 QFDGWEISGFEGKDDAGYVINLSKDTFFIKPV



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 96.6376 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGEVSELKPHRVTVIQNG.....ATVLDKNNISKSTTNPNK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	799	100.0	773	4	AAB48343 S. pneumo
2	799	100.0	2120	3	AAY81710 Streptoco
3	799	100.0	2140	6	ABU01020 S. pneumo
4	799	100.0	2140	6	ABU45746 Protein e
5	799	100.0	2140	8	ADM92113 S pneumon
6	799	100.0	2140	8	ADT50099 S pneumon
7	796	99.6	637	8	ADR94534 Novel S.
8	796	99.6	637	9	AEA58404 Streptoco
9	796	99.6	2138	8	ADK48759 Streptoco
10	615	77.0	117	2	AAW55096 Streptoco
11	615	77.0	117	5	ABF54590 Streptoco
12	615	77.0	117	7	ADC45149 S. pneumo
13	118	14.8	746	5	AAG81779 S. epider
14	118	14.8	778	5	ABP39023 Staphyloc
15	118	14.8	778	6	ADSO6368 Staphyloc
16	110.5	13.8	775	6	ABU42797 Protein e
17	108	13.5	188	9	ADZ79639 P. falcip
18	108	13.5	354	9	ADZ72253 Plasmodiu
19	107.5	13.5	470	8	ADT56185 Plant pol
20	107.5	13.5	484	3	AAG47777 Arabidops
21	106	13.3	647	9	ADZ79635 P. falcip
22	106	13.3	651	8	ADO19012 Amino aci
23	106	13.3	651	8	ADO19010 P. falcip
24	105	13.1	665	3	AAB18278 Plasmodiu

25	105	13.1	665	7	ABO23606	Abc23606 Plasmodiu
26	103.5	13.0	169	9	ADZ79634	P. falcip
27	103	12.9	707	6	ABU25018	Protein e
28	101.5	12.7	564	4	ABB61977	Drosophil
29	97.5	12.2	1184	6	ABU25330	Protein e
30	97	12.1	1791	8	ADP25441	Plasmodiu
31	96.5	12.1	2060	5	AAE20967	Staphyloc
32	96	12.0	639	9	ADW88472	Staphyloc
33	96	12.0	2500	3	AAB18272	Plasmodiu
34	95.5	12.0	635	8	ADW83292	Fibrinoge
35	95.5	12.0	635	8	ADW83292	Streptoco
36	95.5	12.0	643	8	ADV89902	Streptoco
37	95.5	12.0	643	8	ADV81155	Streptoco
38	95	11.9	645	9	ADW88460	Staphyloc
39	95	11.9	645	9	ADW88459	Staphyloc
40	95	11.9	645	9	ADW88458	Staphyloc
41	94	11.8	645	9	ADW88454	Staphyloc
42	94	11.8	645	9	ADW88457	Staphyloc
43	94	11.8	645	9	ADW88453	Staphyloc
44	94	11.8	645	9	ADW88441	Staphyloc
45	94	11.8	645	9	ADW88455	Staphyloc

ALIGNMENTS

RESULT 1  
AAB48343  
ID AAB48343 standard; protein; 773 AA.  
XX  
AC AAB48343;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE S. pneumoniae Spi30 polypeptide.  
XX  
KW Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN WO200076540-A2.  
XX  
PD 21-DEC-2000.  
XX  
PF 09-JUN-2000; 2000WO-US015925.  
XX  
PR 10-JUN-1999; 99US-0138453P.  
XX  
(MEDI-) MED IMMUNE INC.  
PI Adamou JE, Choi GH;  
XX  
XX WPI; 2001-112197/12.  
DR N-PSDB; AAC84742.  
XX  
New vaccines comprising Spi28 or Spi30 polypeptides, for treating and preventing pneumococcal infections, particularly infections caused by Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or blood infections.  
Claim 8; Page 51-54; 54pp; English.  
The invention relates to novel immunogenic polypeptides, Spi28 and Spi30 from S. pneumoniae. Vaccines comprising the polypeptides are useful for the treatment and prevention of pneumococcal infections, particularly infections caused by Streptococcus, such as otitis media, nasopharyngeal, bronchial, lung or blood infections. The antigens are used as immunogenic agents to stimulate an immune response. The antisera and antibodies may also be used in diagnosing and treating pneumococcal infections. Recombinant polypeptides serve as a mechanism for stimulating production of antibodies for use in passive immunotherapy, diagnostic reagents, and

CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Spl30 polypeptide  
XX  
SQ Sequence 773 AA;

Query Match 100.0%; Score 799; DB 4; Length 773;  
Best Local Similarity 100.0%; Pred. No. 6.1e-71;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTTTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 620 DTGEVSELKPHRVTTTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 679  
QY 61 DAGYVINLSKDTFIKPVFKKIEBKKEENKPTFDVSKKONPVNHSQLNESHKEDLQR 120  
DB 680 DAGYVINLSKDTFIKPVFKKIEBKKEENKPTFDVSKKONPVNHSQLNESHKEDLQR 739  
QY 121 EEHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 154  
DB 740 EEHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 773

RESULT 2  
AAY81710  
ID AAY81710 standard; protein; 2120 AA.  
XX  
AC AAY81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PF 27-JUL-1999; 99WO-GB002452.  
XX  
PR 27-JUL-1998; 98GB-00016336.  
PR 19-MAR-1999; 99US-0125329P.  
XX  
PA (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
DR WPI; 2000-195301/17.  
DR N-PSDB; AAZ91806.  
XX  
PT Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections.  
XX  
PS Claim 2; Page 41-42; 76pp; English.  
XX  
CC This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis  
XX  
SQ Sequence 2120 AA;

Query Match 100.0%; Score 799; DB 3; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 2.5e-70;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTTTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 1933 DTGEVSELKPHRVTTTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 1992  
QY 61 DAGYVINLSKDTFIKPVFKKIEBKKEENKPTFDVSKKONPVNHSQLNESHKEDLQR 120  
DB 1993 DAGYVINLSKDTFIKPVFKKIEBKKEENKPTFDVSKKONPVNHSQLNESHKEDLQR 2052  
QY 121 EEHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 154  
DB 2053 EEHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 2086

RESULT 3  
ABU01020  
ID ABU01020 standard; protein; 2140 AA.  
XX  
AC ABU01020;  
XX  
DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX  
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KW gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
FN WO200277021-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-IB002163.  
XX  
PR 27-MAR-2001; 2001GB-00007658.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Masignani V, Tettelin H, Fraser C;  
XX  
DR WPI; 2003-040579/03.  
DR N-PSDB; ABX06302.  
XX  
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
PS Claim 1; SEQ ID NO 1180; 56pp; English.  
XX  
CC The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target

CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 799; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-70;  
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTGEVSELKPHRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEKK 60  
 DB 1953 DTGEVSELKPHRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEKK 2012  
 QY 61 DAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQR 120  
 DB 2013 DAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQR 2072  
 QY 121 EEHSQKSDSTKQVATVLDKNNISSKSTTNNPNK 154  
 DB 2073 EEHSQKSDSTKQVATVLDKNNISSKSTTNNPNK 2106

RESULT 4  
 ABU45746  
 ID ABU45746 standard; protein; 2140 AA.

XX AC ABU45746;  
 XX 19-JUN-2003 (first entry)  
 XX Protein encoded by Prokaryotic essential gene #31273.  
 XX Anti-ense; prokaryotic essential gene; cell proliferation; drug design.  
 XX Streptococcus pneumoniae.  
 XX WO200277183-A2.  
 XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 XX 06-SEP-2001; 2001US-00948993.  
 XX 25-OCT-2001; 2001US-0342923P.  
 XX 08-FEB-2002; 2002US-00072851.  
 XX 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 XX N-PSDB; ACA49616.

XX  
 PT  
 PT  
 PT  
 PS  
 XX

New antisense nucleic acids, useful for identifying proteins or screening  
 for homologous nucleic acids required for cellular proliferation to  
 isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 73670; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of  
 the 6213 antisense sequences given in the specification where expression  
 of the nucleic acid inhibits proliferation of a cell. Also included are:  
 (1) a vector comprising a promoter operably linked to the nucleic acid  
 encoding a polypeptide whose expression is inhibited by the antisense  
 nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 polypeptide or its fragment whose expression is inhibited by the  
 antisense nucleic acid; (4) an antibody capable of specifically binding  
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 proliferation or the activity of a gene in an operon required for  
 proliferation; (7) identifying a compound that influences the activity of  
 the gene product or that has an activity against a biological pathway  
 required for proliferation, or that inhibits cellular proliferation; (8)  
 identifying a gene required for cellular proliferation or the biological  
 pathway in which a proliferation-required gene or its gene product lies  
 or a gene on which the test compound that inhibits proliferation of an  
 organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 compound's activity; (11) a culture comprising strains in which the gene  
 product is overexpressed or underexpressed; (12) determining the extent  
 to which each of the strains is present in a culture or collection of  
 strains; or (13) identifying the target of a compound that inhibits the  
 proliferation of an organism. The antisense nucleic acids are useful for  
 identifying proteins or screening for homologous nucleic acids required  
 for cellular proliferation to isolate candidate molecules for rational  
 drug discovery programs, or for screening homologous nucleic acids  
 required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
*K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 the target prokaryotic essential genes. Note: The sequence data for this  
 patent did not form part of the printed specification, but was obtained  
 in electronic format directly from WIPO at  
 ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 2140 AA;

Query Match 100.0%; Score 799; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-70;  
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DTGEVSELKPHRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEKK 60  
 DB 1953 DTGEVSELKPHRVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEKK 2012  
 QY 61 DAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQR 120  
 DB 2013 DAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQR 2072  
 QY 121 EEHSQKSDSTKQVATVLDKNNISSKSTTNNPNK 154  
 DB 2073 EEHSQKSDSTKQVATVLDKNNISSKSTTNNPNK 2106

RESULT 5  
 ADM92113  
 ID ADM92113 standard; protein; 2140 AA.

XX AC ADM92113;  
 XX 03-JUN-2004 (first entry)

XX S pneumoniae antigenic protein sequence SeqID310.  
 XX antibacterial; gene therapy; Streptococcus pneumoniae infection;  
 XX antigenic.  
 XX Streptococcus pneumoniae.

PN WO2004020609-A2.  
XX  
XX PD 11-MAR-2004.  
XX PF 02-SEP-2003; 2003WO-US027401.  
XX PR 30-AUG-2002; 2002US-0407082P.  
XX PA (TUFT ) UNIV TUFTS.  
XX PI Camilli A, Hava DL;  
XX WPI; 2004-239189/22.  
XX N-PSDB; ADW91876.  
XX  
PT New Streptococcus pneumoniae nucleic acid molecules, useful for  
PT diagnosing, treating and preventing active infections of Streptococcus  
PT pneumoniae.  
XX  
PS Claim 27; SEQ ID NO 310; 123pp; English.  
XX  
CC This invention relates to novel isolated Streptococcus pneumoniae nucleic  
CC acid molecules and the antigenic polypeptides encoded by them. The  
CC invention may be useful for the production of compounds with an  
CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
CC compositions and methods disclosed are useful for treating Streptococcus  
CC pneumoniae infection. The present sequence is that of an S pneumoniae  
CC protein of the invention.  
XX  
XX SQ Sequence 2140 AA;  
  
Query Match 100.0%; Score 799; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.5e-70; Indels 0; Gaps 0;  
Matches 154; Conservative 0; Mismatches 0;  
  
QY 1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 1953 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 2012  
  
QY 61 DAGVINLSKDTFKIPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 120  
DB 2013 DAGVINLSKDTFKIPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 2072  
  
QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
DB 2073 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2106  
  
RESULT 6  
ADT50099  
ID ADT50099 standard; protein; 2140 AA.  
XX  
AC ADT50099;  
XX  
DT 13-JAN-2005 (first entry)  
XX  
DE S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
XX  
KW hyperimmune serum reactive antigen; antibacterial; vaccine;  
KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
KW sepsis; meningitis.  
XX  
OS Streptococcus pneumoniae TIGR4.  
XX  
PN WO2004092209-A2.  
XX  
XX PD 28-OCT-2004.  
XX  
XX PF 15-APR-2004; 2004WO-EP003984.  
XX  
XX PR 15-APR-2003; 2003EP-00450087.  
XX  
XX PA (INTE-) INTERCELL AG.

XX Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;  
XX WPI; 2004-758335/74.  
XX N-PSDB; ADT49955.  
XX  
PT New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
PT treating S. pneumoniae infections.  
XX  
PS Disclosure; SEQ ID NO 177; 191pp; English.  
XX  
CC This invention relates to novel nucleic acids encoding hyperimmune serum  
CC reactive antigens, or fragments derived thereof. Specifically, it refers  
CC to antigens selected from peptides and serum reactive epitopes that can  
CC be used in pharmaceutical compositions that exhibit antibacterial  
CC activity. The present invention describes a composition (including the  
CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
CC that is useful for manufacturing a medicament such as a vaccine, which  
CC can be used to treat or prevent bacterial infections, particularly S.  
CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
CC be used for isolating, purifying and/ or identifying an interaction  
CC partner of the hyperimmune serum reactive antigen, as well as for  
CC manufacturing a functional nucleic acid selected from aptamers and  
CC spiegelmers or for manufacturing a functional ribonucleic acid selected  
CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
CC of the invention.  
XX  
XX SQ Sequence 2140 AA;  
  
Query Match 100.0%; Score 799; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.5e-70; Indels 0; Gaps 0;  
Matches 154; Conservative 0; Mismatches 0;  
  
QY 1 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 60  
DB 1953 DTGEVSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK 2012  
  
QY 61 DAGVINLSKDTFKIPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 120  
DB 2013 DAGVINLSKDTFKIPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQR 2072  
  
QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
DB 2073 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2106  
  
RESULT 7  
ADR94534  
ID ADR94534 standard; protein; 637 AA.  
XX  
AC ADR94534;  
XX  
DT 16-DEC-2004 (first entry)  
XX  
DE Novel S. pneumoniae protein sequence, SEQ ID 3169.  
XX  
KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
KW bacterial infection.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US6800744-B1.  
XX  
XX PD 05-OCT-2004.  
XX  
XX PF 30-JUN-1998; 98US-00107433.  
XX  
XX PR 02-JUL-1997; 97US-0051553P.  
XX PR 12-MAY-1998; 98US-0085131P.  
XX





```
XX AC ADK48759;
XX DE 20-MAY-2004 (first entry)
XX DE Streptococcus pneumoniae protein, Seq ID No 5274.
XX KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX OS Streptococcus pneumoniae.
XX PN US6699703-B1.
XX PD 02-MAR-2004.
XX PF 26-MAY-2000; 2000US-00583110.
XX PR 02-JUL-1997; 97US-0051553P.
XX PR 12-MAY-1998; 98US-0085131P.
XX PR 30-JUN-1998; 98US-00107433.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX DR WPI: 2004-212399/20.
XX DR N-PSDB; ADK46098.
XX PT New nucleic acid molecules and polypeptides useful for diagnosing,
XX PT preventing and treating pathological conditions resulting from bacterial
XX PT infection, e.g. Streptococcus pneumoniae infection, and in drug
XX PT screening.
XX PS Disclosure; SEQ ID NO 5274; 301pp; English.
XX CC The invention relates to isolated Streptococcus pneumoniae nucleic acids
XX CC and polypeptides. The nucleic acids and proteins are useful for
XX CC diagnosing, preventing and treating pathological conditions resulting
XX CC from bacterial infection, such as S. pneumoniae infection. These may also
XX CC be used for drug screening procedures. The present sequence represents a
XX CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
XX CC data for this patent did not appear in the printed specification but was
XX CC obtained in electronic format directly from USPTO at
XX CC seqdata.uspto.gov/sequence.html.
XX CC Sequence 2138 AA;
XX CC Query Match 99.6%; Score 796; DB 8; Length 2138;
XX CC Best Local Similarity 99.4%; Pred. No. 5e-70;
XX CC Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 60
DB 1951 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 2010
QY 61 DAGYVNLKSDTFIKVPFKKIEEKEENKPTFDYKKNQPNVHNSQLNESHKREDLQR 120
DB 2011 DAGYVNLKSDTFIKVPFKKIEEKEENKPTFDYKKNQPNVHNSQLNESHKREDLQR 2070
QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
DB 2071 EDHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2104
RESULT 10
ID AAW55096
XX AAW55096 standard; protein; 117 AA.
XX AC AAW55096;
XX DT 02-OCT-1998 (first entry)
XX DE Streptococcus pneumoniae SP0043 protein.
```

```
XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX KW detection; pneumonia; otitis media; meningitis.
XX OS Streptococcus pneumoniae.
XX PN WO9818930-A2.
XX PD 07-MAY-1998.
XX PF 30-OCT-1997; 97WO-US019422.
XX PR 31-OCT-1996; 96US-0029960P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX DR WPI: 1998-272224/24.
XX DR N-PSDB; AAV27357.
XX PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
XX PT - or their epitope-containing fragments, useful in protective or
XX PT therapeutic vaccines, and for diagnosis.
XX PS Claim 11; Page 62; 118pp; English.
XX CC The present sequence represents a protein from Streptococcus pneumoniae.
XX CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX CC can be useful in vaccines for inducing protective antibodies against
XX CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX CC are used to detect Streptococcus infection (by usual hybridisation or
XX CC amplification methods), also for isolating Streptococcus genes or their
XX CC allelic variants. The protein can be used similarly to detect specific
XX CC antibodies in standard immunoassays, especially for diagnosing or
XX CC monitoring infections. Antibodies which bind the protein are used to
XX CC detect corresponding antigens to purify the protein and for passive
XX CC immunisation (optionally coupled to a toxin). Vaccines are administered,
XX CC e.g. by injection, orally or through the skin, typically at 0.01-1000
XX CC (especially 10-300) mu g/ml per dose
XX CC Sequence 117 AA;
XX CC Query Match 77.0%; Score 615; DB 2; Length 117;
XX CC Best Local Similarity 100.0%; Pred. No. 1.3e-53;
XX CC Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 YKGELEKGYQFDGWEISGFEKKDAGYVNLKSDTFIKVPFKKIEEKEENKPTFDYSK 97
DB 1 YKGELEKGYQFDGWEISGFEKKDAGYVNLKSDTFIKVPFKKIEEKEENKPTFDYSK 60
QY 98 KKNPQVNHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154
DB 61 KKNPQVNHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117
RESULT 11
ID ABP54590
XX ABP54590 standard; protein; 117 AA.
XX AC ABP54590;
XX DT 04-SEP-2002 (first entry)
XX DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.
XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX KW antibacterial; Streptococcal infection; detection.
XX OS Streptococcus pneumoniae.
XX PN US2002061545-A1.
```

```

XX PD 23-MAY-2002.
XX PF
XX PP
XX PR 22-JAN-2001; 2001US-00765272.
XX PP
XX PR 30-OCT-1997; 97US-00961083.
XX PP
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2002-479261/51.
XX DR N-PSDB; ABQ84825.
XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX PT and for preventing or attenuating disease caused by Streptococcus
XX PT infection.
XX PS Claim 11; Page 29; 70pp; English.
XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX CC pneumoniae antigens have antibacterial activity and can be used in
XX CC vaccines. The S. pneumoniae antigens can also be used to prevent or
XX CC attenuate a Streptococcal infection in an animal. The polynucleotides
XX CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX CC of S. pneumoniae ORFs (open reading frames) which are used in an example
XX CC from the present invention
XX SQ Sequence 117 AA;
Query Match 77.0%; Score 615; DB 5; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.3e-53;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 YKGELEKGYQPDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 97
DB 1 YKGELEKGYQPDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 60
QY 98 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 154
DB 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117
RESULT 12
ADC45149
ID ADC45149 standard; protein; 117 AA.
XX AC
XX AC ADC45149;
XX DT 18-DEC-2003 (first entry)
XX DE S. pneumoniae antigenic protein SP043.
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX OS Streptococcus pneumoniae.
XX XX US6573082-B1.
XX PN 03-JUN-2003.
XX PD
XX PP 28-MAR-2000; 2000US-00536784.
XX PR 31-OCT-1996; 96US-0029960P.

```

```

PR 30-OCT-1997; 97US-00961083.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2003-764574/72.
XX DR N-PSDB; ADC45148.
XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX PT useful for producing vaccines for prevention or attenuation of infection
XX PT by Streptococcus pneumoniae.
XX PS Example 1; SEQ ID NO 68; 58pp; English.
XX CC The invention relates to an isolated polynucleotide consisting of a
XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX CC antigens. Also included are making a recombinant vector by inserting the
XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at
XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae
XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX CC antigen nucleic acids are useful as probes for use in diagnostic methods
XX CC for detecting S. pneumoniae gene expression. The present sequence
XX CC represents an S. pneumoniae antigenic protein.
XX SQ Sequence 117 AA;
Query Match 77.0%; Score 615; DB 7; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.3e-53;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 38 YKGELEKGYQPDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 97
DB 1 YKGELEKGYQPDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 60
QY 98 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 154
DB 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117
RESULT 13
AAG81779
ID AAG81779 standard; protein; 746 AA.
XX AC
XX AC AAG81779;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.
XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
XX KW endocarditis.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PP 09-NOV-2000; 2000WO-US030782.
XX PR 09-NOV-1999; 99US-0164258P.
XX PA (GLAX ) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX XX WPI; 2001-316495/33.
XX DR N-PSDB; AAH52629.

```

XX Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
PT useful for vaccinating against infections, e.g. endocarditis.  
XX  
PS Claim 18; Page 208; 2189pp; English.  
XX  
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
(II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce host cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX  
SQ Sequence 746 AA;

Query Match 14.8%; Score 118; DB 4; Length 746;  
Best Local Similarity 27.2%; Pred. No. 0.014;  
Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
Qy 1 DTGEVSELKPHRVTVTTIONGKEMSSITIVSEEDFILPVYK-----GELEKGYQPDGW--- 51  
Db 586 DSVNAQSLKP-----ITINGKQIKQSVKSGTKVLPKSKVLMTDGELTMP-DMTGWTKE 640  
Qy 52 EISGFE-----GKDGAGYVIN--LSKDTFIKPVFKKIEKKEENKPTDVS-----KK 98  
Db 641 DVLAFEDTLKLVSTKNGFVTNQISKGQIIK-----NKKIEVLSAEDT 687  
Qy 99 KDNQVNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISSKSTNN 151  
Db 688 DDQEKTEDESSDNKSKKDADEHSNTSSSTKN-----DKSNADSKNDSDD 734

RESULT 14  
ABP39023  
ID ABP39023 standard; protein; 778 AA.  
XX  
AC ABP39023;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.  
XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-00134001.  
XX  
PR 14-AUG-1997; 97US-0055779P.  
PR 08-NOV-1997; 97US-0064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.  
DR N-PSDB; ABN91568.  
XX  
PT Novel isolated nucleic acid encoding a *Staphylococcus epidermis*  
PT polypeptide, useful for diagnosing and treating bacterial infections.  
XX  
PS Disclosure; SEQ ID NO 3868; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences can  
CC also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life cycle  
CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from the USPTO web site  
XX  
SQ Sequence 778 AA;  
Query Match 14.8%; Score 118; DB 5; Length 778;  
Best Local Similarity 27.2%; Pred. No. 0.015;  
Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
Qy 1 DTGEVSELKPHRVTVTTIONGKEMSSITIVSEEDFILPVYK-----GELEKGYQPDGW--- 51  
Db 618 DSVNAQSLKP-----ITINGKQIKQSVKSGTKVLPKSKVLMTDGELTMP-DMTGWTKE 672  
Qy 52 EISGFE-----GKDGAGYVIN--LSKDTFIKPVFKKIEKKEENKPTDVS-----KK 98  
Db 673 DVLAFEDTLKLVSTKNGFVTNQISKGQIIK-----NKKIEVLSAEDT 719  
Qy 99 KDNQVNHSQLNESHKEDLQREHSHQSDSTKDVATVLDKNNISSKSTNN 151  
Db 720 DDQEKTEDESSDNKSKKDADEHSNTSSSTKN-----DKSNADSKNDSDD 766  
RESULT 15  
ADS06368  
ID ADS06368 standard; protein; 778 AA.  
XX  
AC ADS06368;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Staphylococcus epidermis polypeptide seqid 5663.  
XX  
KW antibacterial; vaccine; antitense therapy; *Staphylococcus epidermidis*;  
KW recombinant expression vector; infection; computer readable medium;  
KW computer based system.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US2004147734-A1.  
XX  
PD 29-JUL-2004.  
XX  
PF 01-DEC-2003; 2003US-00724972.  
XX  
PR 08-NOV-1997; 97US-0064964P.  
PR 13-AUG-1998; 98US-00134001.  
PR 29-NOV-1999; 99US-00450969.  
XX  
PA (DOUC/) DOUCETTE-STAMM L.  
PA (BUSH/) BUSH D.  
XX  
PI Doucette-Stamm L, Bush D;  
XX  
DR WPI; 2004-580138/56.  
DR N-PSDB; ADS02596.  
XX

PT New isolated polypeptide and encoding nucleic acid derived from  
 PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or  
 XX treating an S. epidermidis bacterial infection.  
 PS Claim 17; SEQ ID NO 5663; 741pp; English.

XX The invention describes an isolated nucleic acid comprising a nucleotide  
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
 CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any  
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
 CC given in the specification. Also described are: a recombinant expression  
 CC vector; a cell comprising a recombinant expression vector of (1);  
 CC producing an S. epidermidis polypeptide; an isolated nucleic acid  
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
 CC vaccine composition for prevention or treatment of an S. epidermidis  
 CC infection, comprising a nucleic acid cited above and a carrier; treating  
 CC a subject for S. epidermidis infection; a recombinant or substantially  
 CC pure preparation of an S. epidermidis polypeptide or its fragment; a  
 CC vaccine composition for prevention or treatment of an S. epidermidis  
 CC infection; detecting the presence of a Staphylococcus nucleic acid in a  
 CC sample; a computer readable medium having recorded in it the nucleotide  
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
 CC system for identifying fragments of the Staphylococcus genome of  
 CC commercial importance; a computer based system for identifying fragments  
 CC of the Staphylococcus plasmids of commercial importance; identifying  
 CC commercially important nucleic acid fragments of the Staphylococcus  
 CC genome and/or plasmids; and identifying an expression modulating fragment  
 CC of the Staphylococcus genome and/or plasmids. The methods and  
 CC compositions of the present invention are useful for the diagnosis,  
 CC prevention and/or treatment of an Staphylococcus epidermidis bacterial  
 CC infection. This is the amino acid sequence of a S. epidermis protein of  
 XX the invention.

XX Sequence 778 AA;

Query Match 14.8%; Score 118; DB 8; Length 778;  
 Best Local Similarity 27.2%; Pred. No. 0.015; 57; Indels 46; Gaps 9;  
 Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
 QY 1 DTGEVSELKPHRVTTIQNGKMSSTIVSEEDFILPVYK-----GLEKGYQFDGW--- 51  
 Db 618 DSVNAQSLKP-----ITIGNKQIKQOSVKSCTKVLPHSKVLMVMTDGLTMP-DMTGWTKE 672  
 QY 52 EISGFE-----GKKDAGYVIN--LSKDTPIKPVFKIEKKEENKPTFDV-----KK 98  
 Db 673 DVLAPEDLTKIKVSTKNGFVTNQISKGQIIK-----NKDKIEVLSLSDT 719  
 QY 99 KNPQVNHSQLNESHKEDLQREHSQKSDSTKDTATVLDKNKISSKSTNN 151  
 Db 720 DDQEKTDSDSNKSKKDKADEHSTNTSSSTKN-----DKSNADSKNDSDD 766

RESULT 16  
 ABU42797  
 ID ABU42797 standard; protein; 775 AA.  
 AC ABU42797;  
 XX 19-JUN-2003 (first entry)  
 XX Protein encoded by Prokaryotic essential gene #28324.  
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX Staphylococcus epidermidis.  
 XX WO200277183-A2.  
 XX 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohleen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA46667.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS Claim 25; SEQ ID NO 70721; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 775 AA;

Query Match 13.8%; Score 110.5; DB 6; Length 775;  
 Best Local Similarity 25.0%; Pred. No. 0.084;  
 Matches 42; Conservative 27; Mismatches 66; Indels 33; Gaps 8;  
 QY 1 DTGEVSELKPHRVTTIQNGKMSSTIVSEEDFILPVYK-----GLEKGYQFDGW--- 51  
 Db 615 DSVNAQSLKP-----ITIGNKQIKQOSVKSCTKVLPHSKVLMVMTDGLTMP-DMTGWTKE 669  
 QY 52 EISGFE-----GKKDAGYVIN--LSKDTPIKPVFK-----KIEKKEENKPTFD 94  
 Db 670 DVLAPEDLTKIKVSTKNGFVTNQISKGQIIKKNKIEVLSLSDTDDDEKTDSDSD 729  
 QY 95 VSKKQNPQVNHSQLNESHKEDLQREHSQKSDSTKDTATVLDKNK 142  
 Db 730 KSKKDDKYDEDNSNASSSSKNEKSNADSKNDSDDSTNETSGS--ERNN 775  
 RESULT 17  
 ADZ79639  
 ID ADZ79639 standard; protein; 188 AA.

```

XX AC ADZ79639;
XX DT 14-JUL-2005 (first entry)
XX DE P. falciparum merozoite surface protein 3, MSP3a to MSP3f fragment.
XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;
XX KW immunotherapy; malaria; antimalarial; vaccine.
XX OS Plasmodium falciparum.
XX PN WO2005040206-A1.
XX PD 06-MAY-2005.
XX PF 22-OCT-2004; 2004WO-EP012910.
XX PR 24-OCT-2003; 2003US-00691672.
XX PA (INSP ) INST PASTEUR.
XX PI Drulhe P;
XX DR WPI; 2005-355821/36.
XX PT Chimeric molecule useful for preparing vaccine composition against
XX PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface
XX PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice
XX PT immunized with molecule.
XX PS Disclosure; SEQ ID NO 7; 79pp; English.
XX CC The invention relates to a chimeric molecule that comprises a glutamate-
XX CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino
XX CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite
XX CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380
XX CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises
XX CC antibodies against both polypeptides in mice immunized with it. Also
XX CC described are: (i) a conjugate comprising the chimeric molecule of the
XX CC invention bound to a solid support, (ii) an immunogenic composition
XX CC comprising the chimeric molecule, the conjugate described above, or a
XX CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine
XX CC against malaria comprising the chimeric molecule, the conjugate described
XX CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in
XX CC association with a suitable vehicle, (iv) use of purified and/or
XX CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a
XX CC medicament against malaria, and (v) a medicament for passive
XX CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP
XX CC antibodies. The chimeric molecule of the invention or a mixture of GLURP
XX CC and MSP3 antigens are useful for the preparation of a vaccine composition
XX CC against malaria. This sequence represents Plasmodium falciparum MSP3a to
XX CC MSP3f fragment. Note: The present sequence given as SEQ ID No:7 in the
XX CC Sequence Listing is not mentioned elsewhere in the specification.
XX SQ Sequence 188 AA;

Query Match 13.5%; Score 108; DB 9; Length 188;
Best Local Similarity 23.2%; Pred. No. 0.021;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 17 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKQDAG-----YVINLS 69
DB 15 VLKAKEASS-----YDYIL-----GWFGGGVPEHKKEENMLSHLYVSSKD 55
QY 70 KDTFIKVPFKIEEKKEE-----ENKPTFDVSKKKONPQVNHSQLNESHKRE 116
DB 56 KENISKENDVDLDEKEEAETEEBELEKNEETSEISEDEEEEEKEEENDKKK 115
QY 117 DLQREHSQKSDTKDVTATVLDKNNISSKSTNN 151
DB 116 EQEKEQSNENNDDKKDMEA-----QNLISKQNNN 145

```

```

RESULT 18
ADZ72253
ID ADZ72253 standard; protein; 354 AA.
XX AC ADZ72253;
XX DT 14-JUL-2005 (first entry)
XX DE Plasmodium falciparum MSP-3-like protein, MSP-3-1 SEQ ID NO: 2.
XX KW Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;
XX KW infection; merozoite surface protein 3-like protein; MSP-3-1; antigen.
XX OS Plasmodium falciparum.
XX PN EP1526178-A1.
XX PD 27-APR-2005.
XX PF 24-OCT-2003; 2003EP-00292673.
XX PR 24-OCT-2003; 2003EP-00292673.
XX PA (INSP ) INST PASTEUR.
XX PI Drulhe P;
XX DR WPI; 2005-323987/34.
XX DR N-PSDB; ADZ72252.
XX PT Novel MSP-3-like family genes located on chromosome 10 of Plasmodium
XX PT falciparum, which encode proteins useful for preparing vaccine
XX PT compositions against malaria.
XX PS Disclosure; SEQ ID NO 2; 137pp; English.
XX CC The present invention relates to the protection against malaria. More
XX CC particularly the invention pertains to a family of MSP-3 (merozoite
XX CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-
XX CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium
XX CC falciparum, highly conserved in P. falciparum strains, simultaneously
XX CC expressed in P. falciparum at the erythrocytic stages and encoding
XX CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at
XX CC their N-terminal extremity and which are located at the merozoite
XX CC surface. The characterization of this gene family enables the definition
XX CC of immunogenic and vaccine compositions against P. falciparum. The
XX CC present sequence is the P. falciparum MSP-3-1 protein.
XX SQ Sequence 354 AA;

Query Match 13.5%; Score 108; DB 9; Length 354;
Best Local Similarity 23.2%; Pred. No. 0.051;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 17 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKQDAG-----YVINLS 69
DB 181 VLKAKEASS-----YDYIL-----GWFGGGVPEHKKEENMLSHLYVSSKD 221
QY 70 KDTFIKVPFKIEEKKEE-----ENKPTFDVSKKKONPQVNHSQLNESHKRE 116
DB 222 KENISKENDVDLDEKEEAETEEBELEKNEETSEISEDEEEEEKEEENDKKK 281
QY 117 DLQREHSQKSDTKDVTATVLDKNNISSKSTNN 151
DB 282 EQEKEQSNENNDDKKDMEA-----QNLISKQNNN 311

RESULT 19
ADT56185
ID ADT56185 standard; protein; 470 AA.
XX AC ADT56185;

```

XX 13-JAN-2005 (first entry)  
XX Plant polypeptide, SEQ ID 6262.  
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX Viridiplantae.  
XX US2004216190-A1.  
XX 28-OCT-2004.  
XX 18-DEC-2003; 2003US-00739930.  
XX 28-APR-2003; 2003US-00424599.  
XX 28-APR-2003; 2003US-00425115.  
XX (KOVA/) KOVALIC D K.  
XX Kovalic DK;  
XX WPI; 2004-757369/74.  
XX New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.  
XX Claim 2; SEQ ID NO 6262; 14pp; English.  
XX The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant protein  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
XX Sequence 470 AA;

Query Match 13.5%; Score 107.5; DB 8; Length 470;  
Best Local Similarity 20.1%; Pred. NO. 0.084;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;  
QY 9 KPHRTVTIIONKEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFE-----GKK 60  
DB 82 RENRVTDVTQVNSGESK-----YVQDLARRIRYDE-EATGSSAQRIDHPNQK 129  
QY 61 DAGVYVNLSDKTFIKPVFKKIEKKKEENKPTFDVSKKKN----- 101  
DB 130 NVGITEKAPENSPIEETSHRVDDNKRINNQNFTAAKSENAVSEVSGADHKRAEVMGK 189  
QY 102 PQVNHSQLNE-----SHRKEDLQREHSQKSDSTKVDTATVLDQKNITSSKSTNNPNK 154  
DB 190 PMENRDQVRQTESAEKSHRKENVTKSEKPRQDEGVGKTEAKDKDRNKKKEKTESINK 248  
RESULT 20  
AAG47777  
ID AAG47777 standard; protein; 484 AA.  
XX AC AAG47777;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127482P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 22-JUL-1999; 99US-0145193P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.



PR 29-OCT-1999; 99US-0162142P.  
Query Match 13.5%; Score 107.5; DB 3; Length 484;  
Best Local Similarity 20.1%; Pred. No. 0.086;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;  
QY 9 KPHRVTVTTQNGKMSSTIVSEDPILPVYKGELEKGYQFGWEISGPE-----GKK 60  
DB 96 RENRVTTDTVQNNNGESK-----YVQDLARRIRYDE-EATGSGSAQRIDHPNQK 143  
QY 61 DAGYVINLSKDTFIKPVFKKIBKBEENKPTFDVSKKDN-----101  
DB 144 NVGITEKAPENSPIETSHRVDDNKINQKNFTAAKSENAVSRVSGADHRAEVMCK 203  
QY 102 POWNHSQLNE-----SHRKEDLQREHSQKSDSTKVDTATVLDKNISSTTTNNPNK 154  
DB 204 PMENRDQVRQTESAEKSHRKENKVNKSEKPRDQGVKTEAKDKRNMKEKESKTESINK 262  
RESULT 21  
ADZ79635  
ID ADZ79635 standard; protein; 647 AA.  
XX AC ADZ79635;  
XX DT 14-JUL-2005 (first entry)  
XX DE P. falciparum GLURP-MSP3 fusion protein.  
XX KW immune stimulation; fusion protein; glutamate-rich protein; GLURP;  
XX KW merozoite surface protein 3; MSP3; immunotherapy; malaria; antimalarial;  
XX KW vaccine.  
XX OS Plasmodium falciparum.  
XX OS Synthetic.  
XX PN WO2005040206-A1.  
XX XX  
XX PD 06-MAY-2005.  
XX PF 22-OCT-2004; 2004WO-EP012910.  
XX PR 24-OCT-2003; 2003US-00691672.  
XX PA (INSP ) INST PASTEUR.  
XX PI Druilhe P;  
XX XX  
XX DR WPI; 2005-355821/36.  
XX DR N-PSDB; ADZ79636.  
XX XX  
XX PT Chimeric molecule useful for preparing vaccine composition against  
XX PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
XX PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
XX PT immunized with molecule.  
XX XX  
XX PS Disclosure; SEQ ID NO 3; 79pp; English.  
XX XX  
XX CC The invention relates to a chimeric molecule that comprises a glutamate-  
XX CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
XX CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
XX CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
XX CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
XX CC antibodies against both polypeptides in mice immunized with it. Also  
XX CC described are: (i) a conjugate comprising the chimeric molecule of the  
XX CC invention bound to a solid support, (ii) an immunogenic composition  
XX CC comprising the chimeric molecule, the conjugate described above, or a  
XX CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
XX CC against malaria comprising the chimeric molecule, the conjugate described  
XX CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
XX CC association with a suitable vehicle, (iv) use of purified and/or  
XX CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
XX CC medicament against malaria, and (v) a medicament for passive

CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents plasmodium falciparum GLURP(27-  
CC 500)-MSP3(212-380) fusion protein.  
XX XX  
SQ Sequence 647 AA;  
Query Match 13.3%; Score 106; DB 9; Length 647;  
Best Local Similarity 22.8%; Pred. No. 0.19;  
Matches 43; Conservative 35; Mismatches 55; Indels 56; Gaps 9;  
QY 7 ELKPHRVTVTTQNGKEM-----SSTIVSEDPILPVYKGELEK-----44  
DB 428 ETVREHEETVQSENPKEADNDGNVSNNSLNENEFV-----ESEKSEHARSKAKEA 481  
QY 45 -GYOPD-GWETSIGF--EGKKDAG-----YVINLSKDTFIKPVFKKIBKBEENKPTFDV 95  
DB 482 SSYDYLGWFFGGVPEHKKENMLSHLYVSSKKENISKENDDVLDK-KEEEAEETEE 540  
QY 96 SKKDNPOVNHSQLN-----ESHRKEDLQREHSQKSDSTKVDTATVLDKN 142  
DB 541 ELEKNEBETESIESEDEEBEKEBENEKKEQKEQSNENNNDQKKMEA-----QN 595  
QY 143 ISSKSTTN 151  
DB 596 LISKQNNN 604  
RESULT 22  
ADZ019012  
ID ADZ019012 standard; protein; 651 AA.  
XX AC ADZ019012;  
XX DT 12-AUG-2004 (first entry)  
XX DE Amino acid sequence for P. falciparum GLURP-MSP3 hybrid.  
XX KW Glutamate-rich protein; GLURP-MSP3 fusion protein;  
XX KW merozoite surface protein 3; malarial vaccine; malaria; immune response;  
XX KW antimalarial; immunostimulant.  
XX OS Plasmodium falciparum.  
XX OS Synthetic.  
XX PN WO2004043488-A1.  
XX PD 27-MAY-2004.  
XX PF 06-NOV-2003; 2003WO-DK000759.  
XX PR 12-NOV-2002; 2002DK-00001741.  
XX PR 11-SEP-2003; 2003DK-00001307.  
XX XX  
XX PA (STAT-) STATENS SERUM INST.  
XX PI Theisen M, Jepsen S;  
XX XX  
XX DR WPI; 2004-411650/38.  
XX XX  
XX PT New antigen based vaccine comprising a fusion protein derived from  
XX PT Plasmodium falciparum Glutamate-rich protein, useful in treating or  
XX PT preventing malaria.  
XX XX  
XX PS Disclosure; Fig 2C; 52pp; English.  
XX XX  
XX CC The present invention relates to a fusion protein comprising Plasmodium  
XX CC falciparum glutamate-rich protein (GLURP) coupled to P. falciparum  
XX CC merozoite surface protein 3 (MSP3). The GLURP-MSP3 fusion protein is  
XX CC useful as an antigen based vaccine against malaria. Also disclosed is the  
XX CC polynucleotide sequence encoding the GLURP-MSP3 fusion protein. The  
XX CC polynucleotide sequence is also useful in preparing a vaccine. The

CC vaccine is useful in treating and preventing malaria and for inducing an  
CC immune response against malaria. The present sequence represents P.  
CC falciparum GLURP-MSP3 hybrid.

XX SQ Sequence 651 AA;  
Query Match 13.3%; Score 106; DB 8; Length 651;  
Best Local Similarity 22.8%; Pred. No. 0.19;  
Matches 43; Conservative 35; Mismatches 55; Indels 56; Gaps 9;  
QY 7 ELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGLEK----- 44  
Db 432 ETVEHEETVSQESNPEKADNDGNVSQNNELNENEFV-----ESEKSEHARSKAKEA 485  
QY 45 -GYQPD-GWEISGP--EGKDGAG-----YVINLSKDTPIKPVFKKIEKKKEENKPTFDV 95  
Db 486 SSYDYLGMWFGGGVPEHKKEENMLSHLYVSSKOKENISKENDVDLDE-KEEAEETEE 544  
QY 96 SKKDNQPVNHSQLN-----ESHKEDLQREHSQKSDSTKQVDTATVLDKNN 142  
Db 545 ELEEKNEETESEISEDEEEEEEEKEEENEKKEQKESQNNNDQKQMEA-----QN 599  
QY 143 ISSKSTNN 151  
Db 600 LISKQNNN 608

RESULT 23  
AD019010  
ID AD019010 standard; protein; 651 AA.  
XX AC AD019010;  
XX DT 12-AUG-2004 (first entry)  
XX DE P. falciparum GLURP-MSP3 fusion protein.  
XX KW Glutamate-rich protein; GLURP-MSP3 fusion protein;  
XX KW merozoite surface protein 3; malarial vaccine; malaria; immune response;  
XX KW antimalarial; immunostimulant.  
XX OS Plasmodium falciparum.  
XX OS Synthetic.  
XX PN WO2004043488-A1.  
XX PD 27-MAY-2004.  
XX PF 06-NOV-2003; 2003WO-DK000759.  
XX PR 12-NOV-2002; 2002DK-00001741.  
XX PR 11-SEP-2003; 2003DK-00001307.  
XX PA (STAT-) STATENS SERUM INST.  
XX PI Theisen M, Jepsen S;  
XX WPI; 2004-411650/38.  
XX DR N-PSDB; AD019011.  
XX PT New antigen based vaccine comprising a fusion protein derived from  
XX PT Plasmodium falciparum Glutamate-rich protein, useful in treating or  
XX PT preventing malaria.

XX PS Claim 5; SEQ ID NO 1; 52pp; English.  
XX CC The present invention relates to a fusion protein comprising Plasmodium  
XX CC falciparum glutamate-rich protein (GLURP) coupled to P. falciparum  
XX CC merozoite surface protein 3 (MSP3). The GLURP-MSP3 fusion protein is  
XX CC useful as an antigen based vaccine against malaria. Also disclosed is the  
XX CC polynucleotide sequence encoding the GLURP-MSP3 fusion protein. The  
XX CC polynucleotide sequence is also useful in preparing a vaccine. The  
XX CC vaccine is useful in treating and preventing malaria and for inducing an

CC immune response against malaria. The present sequence represents P.  
CC falciparum GLURP-MSP3 fusion protein.

XX SQ Sequence 651 AA;  
Query Match 13.3%; Score 106; DB 8; Length 651;  
Best Local Similarity 22.8%; Pred. No. 0.19;  
Matches 43; Conservative 35; Mismatches 55; Indels 56; Gaps 9;  
QY 7 ELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGLEK----- 44  
Db 432 ETVEHEETVSQESNPEKADNDGNVSQNNELNENEFV-----ESEKSEHARSKAKEA 485  
QY 45 -GYQPD-GWEISGP--EGKDGAG-----YVINLSKDTPIKPVFKKIEKKKEENKPTFDV 95  
Db 486 SSYDYLGMWFGGGVPEHKKEENMLSHLYVSSKOKENISKENDVDLDE-KEEAEETEE 544  
QY 96 SKKDNQPVNHSQLN-----ESHKEDLQREHSQKSDSTKQVDTATVLDKNN 142  
Db 545 ELEEKNEETESEISEDEEEEEEEKEEENEKKEQKESQNNNDQKQMEA-----QN 599  
QY 143 ISSKSTNN 151  
Db 600 LISKQNNN 608

RESULT 24  
AAB18278  
ID AAB18278 standard; protein; 665 AA.  
XX AC AAB18278;  
XX DT 07-NOV-2000 (first entry)  
XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:135.  
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
XX KW antimalarial; malaria; protozoacide; infection; insecticide.  
XX OS Plasmodium falciparum.  
XX PN WO200025728-A2.  
XX PD 11-MAY-2000.  
XX PF 05-NOV-1999; 99WO-US026796.  
XX PR 05-NOV-1998; 98US-0107131P.  
XX PA (HOFF/) HOFFMAN S.  
XX PA (CARU/) CARUCCI D.  
XX PA (GARD/) GARDNER M.  
XX PA (VENT/) VENTER J C.  
XX PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX WPI; 2000-365347/31.  
XX PT Proteins encoded by chromosome 2 of the human malarial parasite,  
XX PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
XX PT diagnosis of P.falciparum infection.  
XX PS Disclosure; Page 321-322; 577pp; English.

XX CC The present invention describes proteins and their fragments (I) encoded  
XX CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
XX CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
XX CC vaccines against P. falciparum infection comprising (I) or (II); (1) and  
XX CC (II) are useful for the development of vaccines against P. falciparum  
XX CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to  
XX CC immunogens comprising the sequences of (I), are useful in the detection  
XX CC of infection with P. falciparum. Furthermore, (I) (especially when they  
XX CC are rifins or secreted or membrane proteins) can aid the identification

CC of drugs to treat or prevent *P. falciparum* infection, or they can be used  
 CC to identify drug resistance in *P. falciparum*. Sequencing of the  
 CC Plasmodium chromosome 2 and the subsequent identification of proteins  
 CC encoded by it will help to expand our understanding of parasite biology,  
 CC a process hampered by the complexity of the parasitic lifecycle, and  
 CC provide new targets for vaccine and drug development. Parasite resistance  
 CC to drugs and mosquito resistance to insecticides have led to a resurgence  
 CC of malaria in many parts of the world, and there is a pressing need for  
 CC vaccines and new drugs. AAY70078 to AAA70287 and AAB18144 to AAB18352  
 CC represent nucleotide and protein sequences given in the present  
 CC invention, but which are not specifically mentioned within the  
 CC specification  
 XX  
 SQ Sequence 665 AA;

Query Match 13.1%; Score 105; DB 3; Length 665;  
 Best Local Similarity 24.4%; Pred. No. 0.24;  
 Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;  
 QY 1 DTGEVSELKPHRVT-VTIQNGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI--SGPE 57  
 DB 128 EKNKINKSDLRHONELNLSQSK-----NEQDI-----NKNEKGKQ----DISNSNAE 170  
 QY 58 GKQDAGYVNLKDTPIKPVFKKIEKKK-----ENKPTFD-----VSKKQDNQ 103  
 DB 171 NKKD-----VKEGVKELEKKEKKEKISDDHKVEENKKSDDHKVBEENKKSDDHK 218  
 QY 104 VNHSQLNESHRKEDIQR-BEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
 DB 219 VEENKKSDDHKIEVKVKEHEDEEE-----DKKKEKSENKNKDNK 261

RESULT 25  
 ABO233606  
 ID ABO233606 standard; protein; 665 AA.  
 AC ABO233606;  
 DT 04-SEP-2003 (first entry)  
 XX Plasmodium falciparum outlier protein #3.  
 DE Candidate protein identification; pathogen; anti-infective;  
 KW outlier protein; virulence protein; antigen; drug target protein;  
 KW pathogenic organism; antimicrobial.  
 OS Plasmodium falciparum.  
 XX  
 PN US2003039963-A1.

27-FEB-2003.  
 30-MAR-2001; 2001US-00820843.  
 30-MAR-2001; 2001US-00820843.  
 (BRAH/) BRAHMACHARI S K.  
 (RAMA/) RAMACHANDRAN S.  
 (NAND/) NANDI T.  
 (BHIM/) BHIMARAO C.  
 XX  
 PI Brahmachari SK, Ramachandran S, Nandi T, Bhimarao C;  
 XX WPI; 2003-492159/46.

Identifying candidate proteins useful as anti-infectives involves  
 PT matching outlier protein sequences with protein sequences in databases.  
 PT  
 PS Example 7; Page 91-93; 117pp; English.

CC The present invention relates to a method for identifying candidate  
 CC proteins in pathogens useful as anti-infectives. The invention discloses  
 CC a computational method which involves the calculation of several sequence

CC attributes and their subsequence analysis results in the identification  
 CC of outlier proteins in different pathogens. The method is useful for the  
 CC identification of outlier proteins (e.g. virulence proteins, antigens or  
 CC proteins used as drug targets) in pathogenic organisms. The method of the  
 CC invention provides reproducible results as it does not depend on the  
 CC variable biochemical characterisation of proteins. ABO233600-ABO23617  
 XX represent outlier proteins identified from different pathogenic organisms  
 SQ Sequence 665 AA;

Query Match 13.1%; Score 105; DB 7; Length 665;  
 Best Local Similarity 24.4%; Pred. No. 0.24;  
 Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;  
 QY 1 DTGEVSELKPHRVT-VTIQNGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI--SGPE 57  
 DB 128 EKNKINKSDLRHONELNLSQSK-----NEQDI-----NKNEKGKQ----DISNSNAE 170  
 QY 58 GKQDAGYVNLKDTPIKPVFKKIEKKK-----ENKPTFD-----VSKKQDNQ 103  
 DB 171 NKKD-----VKEGVKELEKKEKKEKISDDHKVEENKKSDDHKVBEENKKSDDHK 218  
 QY 104 VNHSQLNESHRKEDIQR-BEHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
 DB 219 VEENKKSDDHKIEVKVKEHEDEEE-----DKKKEKSENKNKDNK 261

RESULT 26  
 ADZ79634  
 ID ADZ79634 standard; protein; 169 AA.  
 XX  
 AC ADZ79634;  
 DT 14-JUL-2005 (first entry)  
 XX  
 DE *P. falciparum* merozoite surface protein 3, amino acid residues 212-380.  
 KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
 KW immunotherapy; malaria; antimalarial; vaccine.  
 OS Plasmodium falciparum.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..169  
 FT /note= "Amino acid residues 212-380 of MSP3"

WO2005040206-A1.  
 XX  
 PD 06-MAY-2005.  
 XX  
 PF 22-OCT-2004; 2004WO-EP012910.  
 XX  
 PR 24-OCT-2003; 2003US-00691672.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Druilhe P;  
 XX  
 DR WPI; 2005-355821/36.

Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.

PS Claim 2; SEQ ID NO 2; 79pp; English.  
 XX  
 CC The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also

CC described are: (i) a conjugate comprising the chimeric molecule of the  
CC invention bound to a solid support, (ii) an immunogenic composition  
CC comprising the chimeric molecule, the conjugate described above, or a  
CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
CC against malaria comprising the chimeric molecule, the conjugate described  
CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
CC association with a suitable vehicle, (iv) use of purified and/or  
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
CC medicament against malaria, and (v) a medicament for passive  
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents Plasmodium falciparum MSP3  
CC protein (amino acid residues 212-380).  
XX  
SQ Sequence 169 AA;

Query Match 13.0%; Score 103.5; DB 9; Length 169;  
Best Local Similarity 25.2%; Pred. No. 0.051;  
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;  
QY 21 KEMSTIVSEDFILPVYKGELEKGYQPCGWEISGF--EGKKDAG-----YVINLSKDTFF 73  
DB 1 KEASS-----YDYIL-----GWBFGGVPEHKKEENMLSHLYVSSKKKENI 41  
QY 74 IKPVFKIKBEKKEENKPTFDVSKKDNQVNVHSLN-----BSHRKEDLQR 120  
DB 42 SKENDVDLDE-KEEAETETEELKEKNEETETESISEDEEEEEEKKEEKKKEQEK 100  
QY 121 BEHSQKSDSTKDVATVLDKNNISSKSTTN 151  
DB 101 EQSNENNQKDMEA-----QNLISKNNNN 126

RESULT 27  
ID ABU25018  
XX ABU25018 standard; protein; 707 AA.  
AC ABU25018;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #10545.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
OS Clostridium difficile.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342523P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR N-PSDB; ACA28888.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 52942; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: the sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 707 AA;

Query Match 12.9%; Score 103; DB 6; Length 707;  
Best Local Similarity 25.0%; Pred. No. 0.42;  
Matches 39; Conservative 25; Mismatches 56; Indels 36; Gaps 5;  
QY 1 DTGEVSELKPHRVTTTQNGKEMSTIVSEDFILPVYKGELEKGYQPCGWEISGFEKK 60  
DB 504 DIGDVEDKD-----TTDKYDVS---NKEDIIEPENKSKKKAKULFG----- 542  
QY 61 DAGVIVNLSKDTFIKPVFKIKBEKKEEN--KPTFDVSKKDNQVNVHSLNESHKEDL 118  
DB 543 -----FIKDNEEVEQEENLNDISPDIILDKPVENNQVKSEIEIQELKE-I 589  
QY 119 QREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 154  
DB 590 KQEPSPQHIEBERSVKIEKPINNNLDEKVVSSNNEK 625

RESULT 28  
ID ABB61977  
XX ABB61977 standard; protein; 564 AA.  
AC ABB61977;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 12723.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
PN WO200171042-A2.  
XX  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US009231.  
PF  
XX

PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL06080.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 564 AA;  
 SQ  
 Query Match 12.7%; Score 101.5; DB 4; Length 564;  
 Best Local Similarity 24.5%; Pred. No. 0.43; Mismatches 29; Indels 25; Gaps 5;  
 Matches 34; Conservative 29; Mismatches 29; Indels 25; Gaps 5;  
 QY 31 EDFILPVYKGELEKGYQFDG-----EISGFEKGDAGYVI-----NLSKDTFK 75  
 DB 78 ELDLTPLSRSFSK--VFDGWDEHDEHGDHVDQEPSEALDDHDDHDDHDEDEB 135  
 QY 76 PVFKKLEKKEKREKPT-----FVSKKKQNPQVNSHLSHNRKEDLQREHSHQKSDS 129  
 DB 136 PLTELELELEBEPTDEPAADVEEYEDSEENNA--GENITAEAESEEBEDND 193  
 QY 130 TKDVTATVLDKNNISKST 148  
 DB 194 EGTVEATVEATTEATTEAT 212  
 RESULT 29  
 ABU25330  
 ID ABU25330 standard; protein; 1184 AA.  
 XX AC ABU25330;  
 XX 19-JUN-2003 (first entry)  
 XX Protein encoded by Prokaryotic essential gene #10857.  
 DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KW Clostridium difficile.  
 OS WO200277183-A2.  
 XX PN 03-OCT-2002.  
 XX PD 21-MAR-2002; 2002WO-US009107.  
 XX PF 21-MAR-2001; 2001US-00815242.  
 PR 26-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-03429238.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA29200.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 53254; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 1184 AA;  
 SQ  
 Query Match 12.2%; Score 97.5; DB 6; Length 1184;  
 Best Local Similarity 26.3%; Pred. No. 3.1;  
 Matches 46; Conservative 30; Mismatches 70; Indels 29; Gaps 9;  
 QY 4 EVSELKPHRVTVTIQ-NGKEMSSTIV--SEEDP--ILPVYKGELEKGYQFDGWEISGPEG 58  
 DB 240 ELSVNEHRKVIKELNEKEKQKNVVEKKQDNKEVEVLQDVIEKSYDYN-SKGVIS 298  
 QY 59 KKDAGYVNLSDTP-----IKPVFKKIEKKE-----EENKPTFDVSKKKD 100  
 DB 299 KKE--QINLIKRIPTNEISRNLEIKDKELNENKQVIEKLESNK--LSGESELS 354  
 QY 101 NPQVNSHLSHNRKEDLQREHSHQKSDSTKDVATVLD-KNNISSKSTTNNPK 154  
 DB 355 TLQENIKVLEGSQKQKIKLSLNNELKESIIDILNKQKQEFNSKGLSTLNANK 409  
 RESULT 30  
 ADP25441  
 ID ADP25441 standard; protein; 1791 AA.  
 XX AC ADP25441;  
 XX 09-SEP-2004 (first entry)  
 XX Plasmodium falciparum antigen amino acid sequence SEQ ID NO:18.

XX Plasmodium falciparum; malaria parasite; antigen; malaria; immunogenic;  
 KW immune response; cytostatic; anti-HIV; virucide; hepatotropic;  
 KW antibacterial; vaccine; cancer; infectious disease; AIDS; hepatitis;  
 KW bacterial infection.  
 XX Plasmodium falciparum.  
 OS WO2004053086-A2.  
 XX 24-JUN-2004.  
 PD  
 XX 08-DEC-2003; 2003WO-US038966.  
 XX  
 XX 06-DEC-2002; 2002US-0431494P.  
 XX  
 XX (EPIM-) EPIMMUNE INC.  
 PA (USNA ) US SEC OF NAVY.  
 XX  
 XX Sette A, Doolan DL, Carucci DJ, Sidney J, Southwood S;  
 PI WPI; 2004-468856/44.  
 XX  
 XX New isolated and/or purified Plasmodium falciparum polynucleotide  
 PT sequences, useful in inducing an immune response for preventing and/or  
 PT treating cancer and infectious diseases, such as AIDS, hepatitis, and  
 PT bacterial infections.  
 XX  
 XX Claim 22; SEQ ID NO 18; 253pp; English.  
 PS  
 XX The present invention describes an isolated and/or purified Plasmodium  
 CC falciparum (malaria parasite) antigen polynucleotide sequence, encoding  
 CC an immunogenic peptide. Also described: (1) a primer or detection probe  
 CC for hybridisation with a target sequence or the amplicon generated from a  
 CC target sequence comprising a sequence of at least 8-30, 35, 40, 45, 50,  
 CC 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 consecutive nucleotides of any  
 CC of the polynucleotide sequences as described above; (2) a DNA chip  
 CC comprising any of the polynucleotide sequences described above; (3) a  
 CC vector comprising a promoter operably linked to any of the nucleic acid  
 CC sequences described above; (4) a host cell transformed by the vector of  
 CC (3) or the polynucleotide described above; (5) a composition comprising a  
 CC carrier and the polynucleotide described above; (6) a method of inducing  
 CC an immune response in an individual comprising the administration of the  
 CC composition of (5) to induce an immune response; (7) an isolated  
 CC polypeptide comprising any of the amino acid sequences as encoded by the  
 CC polynucleotide described above; (8) a composition comprising a carrier  
 CC and the polypeptide of (7); (9) a method of detecting P. falciparum in  
 CC biological samples, comprising contacting a biological sample with the  
 CC isolated polynucleotide and detecting the hybridisation of the isolated  
 CC polynucleotides with nucleic acids contained in the sample; (10) a method  
 CC for eliciting an immune response in an individual, comprising the  
 CC administration of a composition comprising the polypeptides of (7) to an  
 CC individual to induce an immune response in the individual; (11) an  
 CC antibody that specifically binds to the P. falciparum polypeptide of (7);  
 CC and (12) detecting P. falciparum antigens, comprising contacting a sample  
 CC from a subject with the polypeptide of (7) and detecting the presence of  
 CC an antigen-antibody complex or detecting the stimulation of T-cells in  
 CC the sample. The P. falciparum antigens and immunogenic peptides have  
 CC cytostatic, anti-HIV, virucide, hepatotropic and antibacterial  
 CC activities, and can be used in vaccines. The methods and compositions of  
 CC the present invention are useful for inducing an immune response for the  
 CC prevention and/or treatment of cancer and infectious diseases, such as  
 CC AIDS, hepatitis, and bacterial infections. The present sequence  
 CC represents a P. falciparum antigen amino acid sequence, which is used in  
 CC the exemplification of the present invention.  
 XX  
 SQ Sequence 1791 AA;  
 Query Match 12.1%; Score 97; DB 8; Length 1791;  
 Best Local Similarity 24.0%; Pred. No. 6.1;  
 Matches 50; Conservative 32; Mismatches 70; Indels 56; Gaps 8;  
 QY 1 DTGEVSELKPHRVTVTTIQNG-----KEMSSITVSEEDFILPVYKGELEKGYQFCGWEIS 54

Db 638 NTSYVLSPLHLGDIVDNNIKRKKKKEIKTIIVSDDMFTSPVNIKEVYNYEQERKKEIV 697  
 QY 55 G---FEGKKDAGYVINLSKDTFIKVPFKIEEKKKEENK-----PTF- 93  
 Db 698 GNLSDYDKTKIIFPFIKFTKEGRKK--KKIEKKEKCKENNNNNFLYNDYSSYSSPKYK 755  
 QY 94 -----DVSKKKDNPOVNHSQL-----NESHKK---EDLQREHSQKS 127  
 Db 756 DNENNFVIKYIRERKDFQKFDHNFNFPSFLHYNPMONKKNKKNKKNVRNEIPNYT 815  
 QY 128 DSTKD-VTATVLDKNNISSKSTTNNPNK 154  
 Db 816 SSSKDGVSYNFLSDSLFSSDNEYSSDNE 843

RESULT 31  
 AAE20967  
 ID AAE20967 standard; protein; 2060 AA.  
 XX  
 AC AAE20967;  
 XX  
 DT 01-JUL-2002 (first entry)  
 DE Staphylococcus lugdunensis von Willebrand factor binding protein.  
 XX  
 XX Von Willebrand factor binding protein; vWb; immunogen; antibacterial;  
 KW vaccine; infection.  
 XX  
 OS Staphylococcus lugdunensis.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..47  
 FT Protein /label= Signal\_peptide  
 FT 48..2060  
 FT /note= "Mature von Willebrand factor binding protein"  
 XX  
 FN WO200228892-A1.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 06-APR-2001; 2001WO-SE000766.  
 XX  
 PR 04-OCT-2000; 2000SE-00003573.  
 XX  
 XX (BIOS-) BIOSTAPRO AB.  
 XX  
 XX Guss B, Frykberg L, Jacobsson K, Ahlen J, Nilsson M;  
 XX WPI; 2002-304928/34.  
 XX N-PSDB; AAD33371.  
 DR  
 XX New von Willebrand factor binding protein from Staphylococci, useful for  
 PT determining and treating staphylococcal infection.  
 PT  
 PS Claim 3; Page 35-41; 53pp; English.  
 XX  
 CC The present invention relates to von Willebrand factor binding protein or  
 CC polypeptide (vWb) from Staphylococci. The vWb and immunogens of vWb are  
 CC useful in vaccines to combat infections caused by Staphylococci. The  
 CC invention is also useful for detection of staphylococcal infection and  
 CC purifying von Willebrand factor from a complex solution. The present  
 CC sequence is Staphylococcus lugdunensis vWb protein  
 XX  
 SQ Sequence 2060 AA;  
 Query Match 12.1%; Score 96.5; DB 5; Length 2060;  
 Best Local Similarity 25.7%; Pred. No. 8.3;  
 Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;  
 QY 4 EVSELKPHRVTVTTIQ-NGKEMSSITVSEEDFILPVYKGELEKGYQFCG--WEISGFEKK 60  
 Db 1914 DIDLK---ITIVYDTNGRE-----IVPSRKQLPP-EQFIQDMQYTGHK--- 1955







CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.  
 XX  
 SQ Sequence 635 AA;

Query Match 12.0%; Score 95.5; DB 8; Length 635;  
 Best Local Similarity 24.5%; Pred. No. 2;  
 Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;  
 QY 19 NGKMSSTIVSEEDFILP--VYKGL-----EKGYPD-----GWEISGFEKDKDAGYVI 66  
 DB 313 NTEPTLSYLENKEKFLVNPYKKNLILREEDKYSFEDDEBEFGNLLSYNKLKNEVLVPV 372  
 QY 67 NLSKDTFTKPVFKKIEEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS-- 124  
 DB 373 NITTTILKP-----PEQKKIVED---FNPYSNLDNLEIKKIRLNGSQKQKVEQETKSPT 425  
 QY 125 QKSDSTKDVATVLDKN--NISKSTT 149  
 DB 426 POKETVKEQTEQKVSQNTQVEKSET 452

RESULT 36  
 ADV89902  
 ID ADV89902 standard; protein; 643 AA.  
 AC ADV89902;  
 DT 24-FEB-2005 (first entry)  
 DE Streptococcus agalactiae protein sequence, SEQ ID 2296.  
 KW Antibacterial; Vaccine; bacterial infection.  
 OS Streptococcus agalactiae.  
 XX  
 FN FR2824074-A1.  
 PD 31-OCT-2002.  
 PF 26-APR-2001; 2001FR-00005642.  
 XX  
 PR 26-APR-2001; 2001FR-00005642.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Glaser P, Rueniok C, Chevallier F, Frangeul L, Lalloui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;  
 XX  
 DR WPI; 2004-101891/11.  
 XX  
 PT Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX  
 PS Claim 6; SEQ ID NO 2296; 2687pp; French.  
 XX  
 CC The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;

CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
 CC agalactiae involved in the synthesis of amino acids, cell membranes,  
 CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
 CC phospholipid metabolism, nucleotide metabolism including purines,  
 CC pyrimidines and/or nucleosides, regulatory functions, replication,  
 CC transcription, translation, protein transport, adaptation to atypical  
 CC conditions, sensitivity to medicines and/or analogues, functions related  
 CC to transposons, biosynthesis of cofactors, prosthetic groups and  
 CC transporters, cell membrane proteins and cellular machinery. (I) are  
 CC useful for the detection and/or amplification of nucleic acids.  
 CC Pharmaceutical composition comprising (I) or (II) are useful for  
 CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
 CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
 CC contains 6617 sequence whereas the present patent only contains 2344  
 CC sequences.  
 XX  
 SQ Sequence 643 AA;

Query Match 12.0%; Score 95.5; DB 8; Length 643;  
 Best Local Similarity 24.5%; Pred. No. 2.1;  
 Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;  
 QY 19 NGKMSSTIVSEEDFILP--VYKGL-----EKGYPD-----GWEISGFEKDKDAGYVI 66  
 DB 321 NTEPTLSYLENKEKFLVNPYKKNLILREEDKYSFEDDEBEFGNLLSYNKLKNEVLVPV 380  
 QY 67 NLSKDTFTKPVFKKIEEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS-- 124  
 DB 381 NITTTILKP-----PEQKKIVED---FNPYSNLDNLEIKKIRLNGSQKQKVEQETKSPT 433  
 QY 125 QKSDSTKDVATVLDKN--NISKSTT 149  
 DB 434 POKETVKEQTEQKVSQNTQVEKSET 460

RESULT 37  
 ADV81155  
 ID ADV81155 standard; protein; 643 AA.  
 AC ADV81155;  
 DT 24-FEB-2005 (first entry)  
 DE Streptococcus agalactiae protein, SEQ ID 2296.  
 KW Antibacterial; vaccine; bacterial infection.  
 OS Streptococcus agalactiae.  
 XX  
 FN WO200292818-A2.  
 PD 21-NOV-2002.  
 PF 26-APR-2002; 2002WO-IB003059.  
 XX  
 PR 26-APR-2001; 2001FR-00005642.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Glaser P, Rueniok C, Chevallier F, Frangeul L, Lalloui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX  
 DR WPI; 2004-101891/11.  
 XX  
 PT Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX  
 PS Claim 6; SEQ ID NO 2296; 439pp; French.  
 XX  
 CC The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and

CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of *S. agalactiae* involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial *S. agalactiae* infection. The complete  
 CC genome of *Streptococcus agalactiae* is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.

XX  
 SQ Sequence 643 AA;

Query Match 12.0%; Score 95.5; DB 8; Length 643;  
 Best Local Similarity 24.5%; Pred. No. 2.1;  
 Matches 36; Conservative 33; Mismatches 55; Indels 23; Gaps 7;  
 QY 19 NKGKMSSTIVSEDFILP--VYKGELE-----EKGYQFD-----GWEISGFEGKDKAGYVI 66  
 DB 321 NTEPLTSYLENKEKFLVPNIPYKNKILILREEDKYSFEDDEBEFGNELLSSYNKLNKNEVLVPV 380  
 QY 67 NLSKDTFIKVPKKTIEEKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHS-- 124  
 DB 381 NITTTILKLP-----FEQKKIVED---FNPYSNLDNLEIKIRLNGSQKQVEQEKTSP 433  
 QY 125 QKSDSTKQDTATVLDKN--NISKSTT 149  
 DB 434 PQKEIVKEQTEQKVSNGTOVEKKSET 460

RESULT 38

ID ADW88460  
 XX ADW88460 standard; protein; 645 AA.  
 AC ADW88460;  
 XX  
 XX 21-APR-2005 (first entry)  
 DT  
 DE Staphylococcus aureus hybrid ORF0657n polypeptide.  
 XX  
 XX ORF0657n; vaccine; antibacterial; protein engineering;  
 KW Staphylococcus aureus infection; mutin.  
 XX  
 XX Staphylococcus aureus.  
 OS Synthetic.  
 XX  
 XX WO2005009378-A2.  
 FN  
 XX  
 XX 03-FEB-2005.  
 PD  
 XX  
 XX 22-JUL-2004; 2004WO-US023522.  
 PP  
 XX  
 XX 24-JUL-2003; 2003US-0489840P.  
 PR  
 XX  
 XX (MERI ) MERCK & CO INC.  
 PA  
 XX  
 XX Anderson AS, Kuklin N, Jansen KU;  
 PI  
 XX  
 XX WPI; 2005-123069/13.  
 DR

XX Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
 PT useful for inducing protective immune response in humans against  
 PT Staphylococcus aureus infection.

XX Claim 7; SEQ ID NO 29; 84pp; English.

XX The present sequence is that of a Staphylococcus aureus protein ORF0657n  
 CC hybrid polypeptide. This is an example of claimed hybrid polypeptide

CC immunogens of the invention ADW88439-ADW88474 that comprise a modified *S.*  
 CC aureus ORF0657n sequence ADW88433-ADW88438 containing amino acid  
 CC substitutions that increase sequence similarity to ORF0190 ADW88432. The  
 CC hybrid polypeptides contain one or more epitopes for ORF0657n and  
 CC ORF0190. They were designed by taking into account the similarity and  
 CC differences between native ORF0657n and ORF0190 protein sequences. The  
 CC invention also provides nucleic acids encoding these hybrid polypeptides,  
 CC and a method for evaluating the ability of an immunogen to produce a  
 CC protective immune response against *Staphylococcus* infection using an  
 CC animal (mouse or rat) model. The hybrid polypeptides having therapeutic  
 CC and diagnostic applications, such as being used to provide protective  
 CC immunity against *S. aureus* infection, being used to generate antibodies  
 CC to detect the presence of *S. aureus*, and being used to generate  
 CC therapeutic antibodies that target *S. aureus*.

XX  
 SQ Sequence 645 AA;

Query Match 11.9%; Score 95; DB 9; Length 645;  
 Best Local Similarity 24.5%; Pred. No. 2.3;  
 Matches 49; Conservative 20; Mismatches 59; Indels 72; Gaps 10;  
 QY 16 TIQNGKMSSTIVSEDFILPVYKGELE-KGYQF-----DGWEISGFEGKDKAGYVINL 68  
 DB 355 SVENSESMOTFVEH-----PIKTGILNGKKYVMYMTNDDYKDFWVEGKR-----VRTI 405  
 QY 69 SK-----DTFIKPVFKIE-----EKKEENKPTFDVSKKK 99  
 DB 406 SKDPKNTRIIIPYVEGKALYDAIVKVVVTIDYDQYHVRIVDKINTKANTDKSNKK 465  
 QY 100 D-----NPOVNHSQLNESHKEDLQ-----REEHSOKSDSTKQDT- 134  
 DB 466 EQQNSAKKEATPATPSKPTSPVEKESQKDSQKDDNKQLPSVEKENDASSESQKDKTP 525  
 QY 135 ATVLDRKNISKSTNNPNK 154  
 DB 526 ATKPTKGEVSSSTT--PTK 543

RESULT 39

ID ADW88459  
 XX ADW88459 standard; protein; 645 AA.  
 AC ADW88459;  
 XX  
 XX 21-APR-2005 (first entry)  
 DT  
 DE Staphylococcus aureus hybrid ORF0657n polypeptide.  
 XX  
 XX ORF0657n; vaccine; antibacterial; protein engineering;  
 KW Staphylococcus aureus infection; mutin.  
 XX  
 XX Staphylococcus aureus.  
 OS Synthetic.  
 XX  
 XX WO2005009378-A2.  
 FN  
 XX  
 XX 03-FEB-2005.  
 PD  
 XX  
 XX 22-JUL-2004; 2004WO-US023522.  
 PP  
 XX  
 XX 24-JUL-2003; 2003US-0489840P.  
 PR  
 XX  
 XX (MERI ) MERCK & CO INC.  
 PA  
 XX  
 XX Anderson AS, Kuklin N, Jansen KU;  
 PI  
 XX  
 XX WPI; 2005-123069/13.  
 DR  
 XX Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
 PT useful for inducing protective immune response in humans against  
 PT Staphylococcus aureus infection.

XX Claim 7; SEQ ID NO 28; 84pp; English.



**THIS PAGE BLANK (uspto)**

GenCore version 5.1.7

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 99.6107 Seconds  
(without alignments)  
1161.598 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848

Sequence: 1 TVVKEFILNKDTGEVSELKP.....ATVLDKNNISSKSTTNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	848	100.0	2119	2	Q9AHT5_STRPN	Q9Aht5 streptococc
2	848	100.0	2140	2	Q97RY6_STRPN	Q97ry6 streptococc
3	845	99.6	2144	2	Q8DQP7_STRR6	Q8dqp7 streptococc
4	844	99.5	2144	2	Q9S4M8_STRPN	Q9s4m8 streptococc
5	133.5	15.7	300	2	Q4XUI6_PLACH	Q4xui6 plasmodium
6	119	14.0	775	2	Q8CPK8_STAEP	Q8cpk8 staphylococ
7	113.5	13.4	296	2	Q5OLX8_ENTHI	Q5olx8 entamoeba h
8	112.5	13.3	361	2	Q95P15_PLAFA	Q95p15 plasmodium
9	112.5	13.3	379	2	Q9U6C4_PLAFA	Q9u6c4 plasmodium
10	111.5	13.1	346	2	Q9U0G0_PLARE	Q9u0g0 plasmodium
11	111.5	13.1	354	2	Q25995_PLAFA	Q25995 plasmodium
12	111.5	13.1	354	2	Q8LJ55_PLAF7	Q8lj55 plasmodium
13	111.5	13.1	379	2	Q25706_PLAFA	Q25706 plasmodium
14	111.5	13.1	775	2	Q5HQ11_STAEP	Q5hq11 staphylococ
15	111	13.1	1038	2	Q90784_CHICK	Q90784 gallus gall
16	110.5	13.0	379	2	Q25705_PLAFA	Q25705 plasmodium
17	110	13.0	500	2	Q6EGL7_PASTE	Q6egl7 paramecium
18	110	13.0	829	2	Q8L5F3_PLAF7	Q8l5f3 plasmodium
19	109.5	12.9	384	2	Q5OVJ0_ENTHI	Q5ovj0 entamoeba h
20	109.5	12.9	609	2	Q8I2K8_PLAF7	Q8i2k8 plasmodium
21	109	12.9	380	2	Q26019_PLAFA	Q26019 plasmodium
22	109	12.9	3008	2	Q8I436_PLAF7	Q8i436 plasmodium
23	108.5	12.8	600	2	Q77355_PLAF7	Q77355 plasmodium
24	108.5	12.8	1069	2	Q5I2T7_ENTHI	Q5i2t7 entamoeba h
25	108	12.7	467	2	Q59PE2_CANAL	Q59pe2 candida alb
26	108	12.7	467	2	Q59PL2_CANAL	Q59pl2 candida alb
27	108	12.7	662	2	Q4YMU4_PLARE	Q4ymu4 plasmodium
28	107.5	12.7	470	2	Q9FKJ9_ARATH	Q9fkj9 arabidopsis
29	107.5	12.7	1455	2	Q640L5_MOUSE	Q640l5 mus musculu
30	106.5	12.6	374	2	Q5V9M0_PLAKN	Q5v9m0 plasmodium
31	106.5	12.6	1015	2	Q5W5T1_TETPY	Q5w5t1 tetrahymena

32 106 12.5 616 2 Q6BRW2\_DEBHA Q6brw2 debaryomyce  
33 106 12.5 951 2 Q96229\_PLAF7 Q96229 plasmodium  
34 105.5 12.4 605 2 Q7RJ33\_PLAYO Q7rjc3 plasmodium  
35 105.5 12.4 674 2 Q7RL87\_PLAYO Q7rle7 plasmodium  
36 105 12.4 1550 2 Q54GS1\_DICDI Q54gs1 dictyosteli  
37 104.5 12.3 329 2 Q9NFV9\_PLAFA Q9nfv9 plasmodium  
38 104.5 12.3 736 2 Q4YVY2\_PLABE Q4yvvy2 plasmodium  
39 104 12.3 540 2 Q94CS9\_ARATH Q94cs9 arabidopsis  
40 104 12.3 540 2 Q9SAB4\_ARATH Q9sa84 arabidopsis  
41 103.5 12.2 325 2 Q44016\_DICDI Q44016 dictyosteli  
42 103.5 12.2 393 2 Q7RKU2\_PLAYO Q7rku2 plasmodium  
43 103 12.1 2081 2 Q9LH98\_ARATH Q9lh98 arabidopsis  
44 103 12.1 2223 2 Q5TNJ2\_ANOGA Q5cnj2 anophales g  
45 102.5 12.1 238 2 Q8I226\_PLAF7 Q8i226 plasmodium

## ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN  
ID Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5\_1  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=1179332;  
RX DOI=10.1128/IAI.69.3.1593-1598.2001;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
RA Langermann S., Johnson S., Koenig S.;  
RT "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL; AF291699; AAK19159.1; -, Genomic\_DNA.  
DR HSP; P00782; 2SPT.  
DR MEROPS; S08.064; -.  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
KW Cell wall; Protease.

TIGRFAMS; TIGR01167; LPTXG_anchor; 1.
DR PROSITE; PSS0847; GRAM POS ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Csil wall; Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240426 MW; FA44ADBE2938B334 CRC64;

  

Query Match	100.0%; Score 848; DB 2; Length 2140;
Best Local Similarity	100.0%; Pred No. 2.4e-50;
Matches 164; Conservative	0; Mismatches 0; Indels 0; Gaps 0;

  

QY	1	TTVKEPFILNKDTGEVSELKPHRVTV--IQNGKMSSTIVSEEDPILPVYKGELEKGQFDG	60
Db	1943	TTVKEPFILNKDTGEVSELKPHRVTV--IQNGKMSSTIVSEEDPILPVYKGELEKGQFDG	2002
QY	61	WEISGEGKKDAGYVINLSKDTFKIVPKIKBEKEEENKPTPDVSKKKNPOVNHSQLN	120
Db	2003	WEISGEGKKDAGYVINLSKDTFKIVPKIKBEKEEENKPTPDVSKKKNPOVNHSQLN	2062
QY	121	ESHRKEDLQREHSHSQSDSKTDKVTA-VLDKNNISSKSTTNPNK	164
Db	2063	ESHRKEDLQREHSHSQSDSKTDKVTA-VLDKNNISSKSTTNPNK	2106

  

RESULT 3

Q8DQP7\_STRR6

ID	Q8DQP7_STRR6 PRELIMINARY;	PRT:	2144 AA.
AC	Q8DQP7		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2003 (TrEMBLrel. 26, Last annotation update)		
DE	Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).		
GN	Name=prtA; OrderedLocusNames=spr0531;		
OS	Streptococcus pneumoniae (strain ATCC BAA-255 / R6).		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
OC	Streptococcus.		
RN	NCBI_TaxID=171101;		
OX	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RX	MSDLINSE=21429245; PubMed=11544234;		
RK	DOI=10.1128/JB.183.19.5709-5717.2001;		
RA	Hoskins J., Albom W.B. Jr., Arnold J., Blaszcak L.C., Burgett S.,		
RA	DeHoff B.S., Batrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,		
RA	Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,		
RA	Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,		
RA	McAlaran S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,		
RA	Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Roake P.,		
RA	Sun P.-M., Winkler M.B., Yang Y., Young-Bellido M., Zhao G.,		
RA	Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,		
RA	Glass J.I.;		
RT	"genome of the bacterium Streptococcus pneumoniae strain R6.";		
RL	J. Bacteriol. 183:5709-5717(2001).		
RM	EWBL; AE008434; AAK99365.1; -; Genomic_DNA.		
DR	PIR; A97942; A97942.		
DR	HSPP; P00782; 2SBT.		
DR	MEROPS; S08.064; -.		
DR	GO; GO:0009286; C:cell surface; IEA.		
DR	GO; GO:0005618; C:cell wall; IEA.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0008233; F:peptidase activity; IEA.		
DR	GO; GO:0042802; F:protein self binding; IEA.		
DR	GO; GO:0004289; F:subtilase activity; IEA.		
DR	GO; GO:0043086; P:negative regulation of enzyme activity; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		
DR	InterPro; IPRO10435; DUFI034.		
DR	InterPro; IPRO01899; Gram_pos_anchor.		
DR	InterPro; IPRO03137; PA.		
DR	InterPro; IPRO00209; Pept_S8_S53.		
DR	InterPro; IPRO10259; Prot_inh_SA.		
DR	InterPro; IPRO01680; WD40.		
DR	Pfam; PF06280; DUFI034; 1.		
DR	Pfam; PF00746; Gram_pos_anchor; 1.		

DR	PRINTS; PRO0723; SUBTILISIN.
DR	TIGRFAMS; TIGR01167; LPXTG anchor; 1.
DR	PROSITE; PSS0847; GRAM POS ANCHORING; 1.
DR	PROSITE; PS00137; SUBTILASE HIS; UNKNOWN_1.
DR	PROSITE; PS00139; SUBTILASE_SER; UNKNOWN_1.
DR	PROSITE; PSQ0678; WD_REPEATS_1; UNKNOWN_1.
FT	Cell wall; Signal.
FT	SIGNAL 1 Potential.
FT	CHAIN 20 2144 cell wall-associated serine proteinase PtaA.
SQ	SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;
Query Match	99.5%; Score 844; DB 2; Length 2144;
Best Local Similarity	98.8%; Pred. No. 4.6e-50;
Matches 162; Conservative	2; Mismatches 0; Indels 0; Gaps 0;
Qy	1 TTWKPEFLNKDTGEVSELKPHRVTVITIQNGKMSSTIVSEEDFILPVYKGELEKGYPDG 60       :       :
Dd	1947 TTWKPEFLNKDTGEVSELKPHRTVTITIQNGKMSSTIVSEEDFILPVYKGELEKGYPDG 2006       :       :
Qy	61 WEISGPEGKKDAGVINLSKDTPIKVPFKIIEBKGEENKPTFDVSKCKDNPOVNHSQLN 120       :       :
Dd	2007 WEISGPEGKKDAGVINLSKDTPIKVPFKIIEBKGEENKPTFDVSKCKDNPOVNHSQLN 2066       :       :
Qy	121 ESHRKEDLQRREHSQKSDTSKVOTATVLDKNNTSSKSTTNPNPK 164       :       :
Dd	2067 ESHRKEDLQRREHSQKSDTSKVOTATVLDKNNTSSKSTTNPNPK 2110       :       :

```

RESULT 5
Q4XUI6_PLACH PRELIMINARY; PRT; 300 AA.
ID Q4XUI6 PLACH PRELIMINARY;
AC Q4XUI6;
DT 13-SEP-2005 (T-EMBLrel. 31, Created)
DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
DE Hypothetical protein (Fragment).
GN ORFNames=PC000286.03.0;
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5825;
RX [1]
RN NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Jause C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.,
RA "A comprehensive survey of the Plasmodium life cycle by genomic,
RA transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC ENBL; CAAJ01003049; CAH79425.1; -; Genomic_DNA.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD0000018; WD40; 3.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS50082; WD_REPEATS_2; 4.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
FT NON TER
FT 1
FT SEQUENCE 300 AA; 34469 MW; 8EDE4E512AFB1945 CRC64;
Query Match 15.7%; Score 133.5; DB 2; Length 300;
Best Local Similarity 26.6%; Pred. No. 0.18;
Matches 42; Conservative 32; Mismatches 51; Indels 33; Gaps 7;
OV 10 KOTGKVSRLKBPHTVTTIONGKMSSTIVSR-----RDFTLPVVGKELRGVDFCS 61

```

```
Db 121 KSTRLLGHQKPVIRHQPSPNGKFIASSPDKSIRIWSGIDGTYLAVYRHHVGPAYKI-AW 179
QY 62 EISGFEKKDAGYVNLKSDTFIK-----PVFKJIEEKE-----BENKPTFDVSKKDD 110
Db 180 SI-----DNNYIVSCSQDSTLKLWRINHLVPLLRKEENAEQTKDSQK-----NEQKE 227
QY 111 NFQVNSQLNESHREKDLQREHSOKSDSTKDVATVL 148
Db 228 NFQ-NDQNDPNDSEAEKKKKEKNDKTKIKILL 264

RESULT 6
Q8CPK8 STAEPP PRELIMINARY; PRT; 775 AA.
AC Q8CPK8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Penicillin-binding protein 1.
GN OrderedLocusNames=SE0856;
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 12228;
RX PubMed=12950922; DOI=10.1046/j.1365-2958.2003.03671.x;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang H.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593 (2003).
DR EMBL; AE016746; AAO04453.1; -; Genomic_DNA.
DR HSSP; F14677; IQME.
DR GO; GO:0008658; F:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR005311; PBP_dimer.
DR InterPro; IPR012338; PBP_tpept_fold.
DR InterPro; IPR001460; Pencil_bind_tpept.
DR Pfam; PF03793; PASTA; 2.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR SMART; SM00740; PASTA; 2.
KW Complete proteome.
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;

Query Match 14.0%; Score 119; DB 2; Length 775;
Best Local Similarity 27.0%; Pred. No. 5.1;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

QY 10 KDTGEVSELKPRHVTVTIQNGKEMSSITVSEDFILPVYK-----GELEKGYQPDGW-- 61
Db 614 EDSVNAQSLKP-----ITIGNHQIKQSVKSTKVLPHSKVLMWTDGELTWP-DMTGWTK 668
QY 62 -SISGFE-----GKKDAGYVIN--LSKDTPIKPVFKIEEKEENKPTFDVS-----K 107
Db 669 EDVLAFEDLTIKVSTKNGFVTSQISKGQIIK-----NKDKIEVLSAED 715
QY 108 KKDNPQVNSQLNESHREKDLQREHSOKSDSTKDVATVLDDKNISKSTTN 161
Db 716 TDDQEKTDSDSNKSKKDAEDHSNTSSSTKN-----DKSNADSKNSDD 763

RESULT 7
Q50LX8 ENTHI
ID Q50LX8 ENTHI PRELIMINARY; PRT; 296 AA.
AC Q50LX8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
```

```
DE Hypothetical protein.
GN ORFNames=657.t00001;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=HM-1:IMSS;
RC PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amadeo P., Rongaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
KW EMBL; AAPB01001439; EAL42595.1; -; Genomic_DNA.
SQ SEQUENCE 236 AA; 33757 MW; 3A5986BB34A7FC3B CRC64;

Query Match 13.4%; Score 113.5; DB 2; Length 296;
Best Local Similarity 26.1%; Pred. No. 4.4;
Matches 41; Conservative 20; Mismatches 63; Indels 33; Gaps 5;

QY 4 KEPLNKDTGEVSELKPRHVTVTIQNGKEMSSITVSEDFILPVYKGELEKGYQPDGW 63
Db 135 KEQETKESGEGSEKKH--DIPTNEGKNKDTTKDND-----KEEKDTWEEG-ES 184
QY 64 SGFGKDGAGVYVNLKDTPIKPVFKIEEKEENKPTFDVSKKNPNQVNSQLNESH 123
Db 185 SKEGGQTN-----ESKKTENPQNNESSNKEEQ----KKEEEK 221

QY 124 RKEDLQREHSOKSDSTKDVATVLDDKNISKSTTN 160
Db 222 KKEEQKKEBEQKKEEQNQEKPNKNEQKENTTN 258

RESULT 8
Q95PI5 PLAPA
ID Q95PI5 PLAPA PRELIMINARY; PRT; 361 AA.
AC Q95PI5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=FVO;
RC MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;
RA Hiseada H., Saul A., Reece J.J., Kennedy M.C., Long C.A., Miller L.H.,
RA Stowers A.W.;
RT "Merozoite surface protein 3 and protection against malaria in Aotus
RT nancymai monkeys.";
RL J. Infect. Dis. 185:657-664(2002).
DR EMBL; AY044180; AAK94780.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
FT CHAIN <1 >361 merozoite surface protein 3.
FT NON_TER 1 1
FT NON_TER 361 361
```



```
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 13.3%; Score 112.5; DB 2; Length 361;
Best Local Similarity 22.4%; Pred. No. 6.4;
Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;

QY 11 DTGEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQ-----57
Db 147 ETGE-----RNSRNFFYTKTK-----YAGKEKDYERAKNAYQKANQAV 187

QY 58 -----FD---GWEISGF--EGKKDAG-----YVNLKSDTFIKPVFKIEEKKEEN 99
Db 188 LKAKERASSYDILGWFEFGGVPEHKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEBA 246

QY 100 KPTFDVSKKQNPQVNHSLNHRKEDLQREHSOKSDSTKQVTVATVLDKN-----151
Db 247 BETESELKEKEEETSEISEDEBEKEEKEEENDKKEKEQKQSNENNNDQKQDME 306

QY 152 --NISKSTNN 161
Db 307 AQNLISKQNNN 318

RESULT 9
Q9U6C4_PLAPA
ID Q9U6C4_PLAPA PRELIMINARY; PRT; 379 AA.
AC Q9U6C4;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Polymorphic antigen.
GN Name:MSP-3;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PCCI/HN;
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188190; AAF04099.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;

Query Match 13.3%; Score 112.5; DB 2; Length 379;
Best Local Similarity 23.6%; Pred. No. 6.8;
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;

QY 19 KPHRVTVTIQNGKEMSSIVSEEDF-----ILPVYKGELEKGYQFD-GWEISGF- 66
Db 171 KPSRLNLSRKTKYAEQV--EKDYERAKNAYQKANQAVLKAKKEASSYDYLGWFEFGGV 228

QY 67 -EGKKDAG-----YVNLKSDTFIKPVFKIEEKKEENKPTFDVSKKQNPQVNHSLN 120
Db 229 PEHKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEBAEETKEEKEEETSEIS 287

QY 121 -----ESHRKEDLQREHSOKSDSTKQVTVATVLDKNISKSTNN 161
Db 288 EDEBEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE 336

RESULT 10
Q9U0G0_PLARE
ID Q9U0G0_PLARE PRELIMINARY; PRT; 346 AA.
AC Q9U0G0;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN Name:mep3;
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
```

```
OX NCBI_TaxID=5854;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
RT Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252286; CAB65754.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
KT Merozoite.
FT NON_TER 1 1
FT NON_TER 346 346
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAF010 CRC64;

Query Match 13.1%; Score 111.5; DB 2; Length 346;
Best Local Similarity 23.9%; Pred. No. 7.2;
Matches 42; Conservative 32; Mismatches 67; Indels 35; Gaps 7;

QY 1 TTVKEFILN-KDTGEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQFD 59
Db 157 TKTKEYAQKAKNAYEKAKNAYQKANQAVLKAKKEASS-----YNYIL----- 197

QY 60 GWEISGF--EGKKDAG-----YVNLKSDTFIKPVFKIEEKKEENKPTFDVSKKQNP 112
Db 198 GWFEFGGVPEHKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEBAEETGELEKKE 256

QY 113 QVNHSLNHRKEDLQREHSOKSDSTKQVTVATV-----LDKNISKSTNN 161
Db 257 BETESELKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE 312

RESULT 11
Q25995_PLAPA
ID Q25995_PLAPA PRELIMINARY; PRT; 354 AA.
AC Q25995;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NF54;
RX MEDLINE=95198774; PubMed=7891748; DOI=10.1016/0166-6851(94)00149-9;
RA McColl D.J., Silva A., Foley M., Kun J.P., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites.";
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NF54;
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L28825; AAC09377.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 13.1%; Score 111.5; DB 2; Length 354;
Best Local Similarity 22.5%; Pred. No. 7.4;
Matches 41; Conservative 34; Mismatches 62; Indels 45; Gaps 7;

QY 1 TTVKEFILN-KDTGEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQFD 59
Db 154 TKTKEYAFKAKNAYEKAKNAYQKANQAVLKAKKEASS-----YDYIL----- 194
```

DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)	
DT	01-MAR-2004 (TReMBLrel. 26, Last annotation update)	
DE	Polymorphic antigen.	
DE	Plasmodium falciparum.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxId=5833;	
RP	[1]	
RP	NUCLEOTIDE SEQUENCE.	
RA	MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;	
RA	McColl D.J., Anders R.F.;	
RT	"Conservation of structural motifs and antigenic diversity in the	
RT	Plasmodium falciparum merozoite surface protein-3 (MSP-3).";	
RL	Mol. Biochem. Parasitol. 90:21-31(1997).	
RL	ENBL; U08852; AAC47832.1; -; Unassigned_DNA.	
DR	InterPro; IPR010784; Merozoite SPAM.	
DR	Pfam; PF07133; Merozoite SPAM; 1	
SQ	SEQUENCE 379 AA; 43302 MW; AAF3D54H1ED91A24 CRC64;	
	Query Match 13.1%; Score 111.5; DB 2; Length 379;	
	Best Local Similarity 23.6%; Pred. No. 7.9;	
	Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps	
Qy	19 KPHRVTVTIQNGKMSSTIVSEDF-----LLPVYKGELEKGVQFD-GWEISFG	
Db	171 KESRLNLSFRKTEAQAQV--EKDYERAKNAVYAKANQAVLKAKBASSYDYILGWFGG	
Qy	67 -GGKKDAG-----YVNLSKDTPTKVPFKKIEBKKEENKPTFDVSKKDDNPQVNHSL	
Db	229 PEHKKEENLHSLVSSKDKENTSKENDVDLDE-KEEEAEETEESELEKKEEETESIE	
Qy	121 -----BSHRKBLQREHSQKSDSTKQVTVATVLDKNNISKSTNN 161	
Db	288 EDEEEEEEKKEENDKKCKEKEQENENNDDQKDEA-----QNLISKNNQNN 336	
RESULT 14		
Q5HQ11_STAEQ		
ID	Q5HQ11_STAEQ PRELIMINARY; PRT; 775 AA.	
AC	Q5HQ11	
DT	10-MAY-2005 (TReMBLrel. 30, Created)	
DT	10-MAY-2005 (TReMBLrel. 30, Last sequence update)	
DT	10-MAY-2005 (TReMBLrel. 30, Last annotation update)	
DE	Penicillin-binding protein 1.	
GN	Names=pbp1; OrderedLocusNames=SERP0746;	
OS	Staphylococcus epidermidis (strain ATCC 35984 / RP62A).	
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
OX	NCBI_TaxId=176279;	
RP	[1]	
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].	
RA	PubMed=15774886; DOI=10.1128/JB.187.7.2426-2438.2005;	
RA	Gill S.R., Fouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T.,	
RA	Ravel J.J., Paulsen I.T., Kolonay J.P., Brinkac L.M., Beanan M.J.,	
RA	Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S.,	
RA	Haft D.H., Vamathevan J.J., Khouri H., Utterback T.R., Lee C.,	
RA	Dimitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H.,	
RA	Hance I.R., Nelson K.E., Fraser C.M.;	
RT	"Insights on evolution of virulence and resistance from the complete	
RT	genome analysis of an early methicillin-resistant Staphylococcus	
RT	aureus strain and a biofilm-producing methicillin-resistant	
RT	Staphylococcus epidermidis strain.;"	
RL	J. Bacteriol. 187:2426-2438(2005).	
DR	ENBL; CP000029; AAW54126.1; -; Genomic_DNA.	
DR	TIGR; SERP0746; -	
DR	GO; GO:008658; F:penicillin binding; IEA.	
DR	GO; GO:009273; F:cell wall biosynthesis (sensu Bacteria); IEA.	
DR	InterPro; IPR005543; PASTA.	
DR	InterPro; IPR005311; PBP dimer.	
DR	InterPro; IPR001460; PenC_bind_tpept.	
DR	Pfam; PF03793; PASTA; 2	
DR	Pfam; PF03717; PBP_dimer; 1.	
DR	Pfam; PF00905; Transpeptidase; 1.	
DR	SMART; SM00740; PASTA; 2.	
KW	Complete proteome.	

```

SQ SEQUENCE 775 AA; 86352 MW; B9395893E0043694 CRC64;

Query Match 13.1%; Score 111.5; DB 2; Length 775;
Best Local Similarity 24.9%; Pred. No. 17;
Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;

Qy 10 KOTGEVSELKPRVTVTIIQNGKEMSSSTIVSEDFILPVYK-----GHELEKGYQPDGW-- 61
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
614 EDSVNAQSLKP---ITIGNGKQIKQOSVKSQGTGKVLPHSKVYMLMTDGLTWP-DMTGWTK 668
Qy 62 -EISGFE-----GKKDAGVYIN--LSKOTFIKPVFK-----KLEEKKEENKPTF 103
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
669 EDVLAFEDLTIKIVSTKNGFVYQWSISKQIINKKDKIEVLSAEATDDDDQKTDSS 728
Qy 104 DVSKKONQVNHSQLNESHKRDQLREHSQKSDSTKDVATVLDKNN 152
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
729 DKSKKQKVDENSNASSSSKVEKSNADSKNDSDDSTNETSGS--ERNN 775

RESULT 15
Q90784_CHICK
ID Q90784_CHICK PRELIMINARY; PRT; 1038 AA.
AC Q90784;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Claustrin.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
RN R1 NCBI
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=94157526; PubMed=7906711;
RA Burg M.A., Cole G.J.;
RT "Claustrin, an anticholinergic neural keratan sulfate proteoglycan, is
RT structurally related to MAP1B.";
RL J. Neurobiol. 25:1-22(1994).
DR EMBL; X67778; CAA47988.1; -, mRNA.
DR F1R; JC5497; JC5497.
DR Ensembl; ENSGALG0000014999; Gallus gallus.
SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;

Query Match 13.1%; Score 111; DB 2; Length 1038;
Best Local Similarity 23.7%; Pred. No. 25;
Matches 44; Conservative 28; Mismatches 58; Indels 56; Gaps 5;

Qy 3 VKGFILNKDTGSEVSELKPRVTVTIIQNGKEMSSSTIVSEDFILPVYK-----GHELEKGYQPDGW-- 61
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
528 VKQAKLQRTDSKESLAPAAKTTTKQDQKRNKKHWSLQSLVQOLEKPKQLSEKKEPT 587
Qy 30 -----GKEMSSSTIVSEDFILPVYKELKGYQPDGWEISGFGKKDAGVYINLSKDTF 83
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
588 VKKEKAVKPKETKITVAEKDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKES 639
Qy 84 IKPVFK-KIEKKEENKPTFDVSKKONQVNHSQLNESHKRDQLREHSQKSDSTKDVATVLDKNN 152
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
640 VKKEVKAPEKKDKEKPKKEVSKGEKPLI---KKEEKPKGEDIKKEVKEKVKGEKK 696
Qy 138 DSTKDV 143
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
697 EAKKEV 702

RESULT 16
Q25705_PLAFA
ID Q25705_PLAFA PRELIMINARY; PRT; 379 AA.
AC Q25705;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

```

DE	Polymorphic antigen.
OS	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_TaxID=5633;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RA	MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RR	McColl D.J., Anders R.P.;
RT	"Conservation of structural motifs and antigenic diversity in the
RT	Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL	Mol. Biochem. Parasitol. 90:21-31(1997).
DR	EMBL; U08851; AAC47831.1; -; Unassigned_DNA.
DR	InterPro; IPK010784; Merozoite_SPAM.
DR	Pfam; PF07133; Merozoite_SPAM; 1.
SQ	SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;
Query Match	13.0%; Score 110.5; DB 2; Length 379;
Best Local Similarity	22.1%; Pred. No. 9.3;
Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9	
Qy	11 DTGSEVLKPHRVTVTTIQNGKMSSTVSBDFILPVYKGLEKYQ----- 57
Dd	167 ETGE-----RNRNNPYTYTKE-----YAGVKDYERAKNAYQKANQAV 207
Qy	58 -----FD--GWETSGF--EGKKDAG-----YVINLSKDTPIKPVPFKIKBEKKBEEN 99
Dd	208 LKAKEASSYDVILGWFGGGVPEHKKEENMLSHLYSSKDKENISKENDVDLDE-KBEEA 266
Qy	100 KPTFDVSKKDNQPNVHSQLN-----ESHKVEDLRQEHSQKSDTKDVTAT 146
Dd	267 EETEELBELBKNEETSISISEDEEEEEBEKEEKKCKEQEKESQNNENNDDKKOMEA- 325
Qy	147 VLDKNISSSKSTTN 161
Dd	326 ----QNLSKNQNN 336
RESULT 17	
Q6BGL7 PARTE	
ID Q6BGL7 PARTE PRELIMINARY; PRT; 500 AA.	
IC Q6BGL7;	
DT 25-OCT-2004 (TrEMBLrel. 28, Created)	
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)	
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)	
DE Hypothetical protein.	
GN OXPNames=PTWB.06c;	
OS Paramacium tetraurelia.	
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;	
OC Paramacium.	
ON NCBI_TaxID=5888;	
RN [1]	
RP NUCLEOTIDE SEQUENCE.	
RC STRAIN=Stock d4-2;	
RX PubMed=15296759; DOI=10.1016/j.cub.2004.07.029;	
RA Zagulska M., Nowak J.K., Le Mouel A., Nowacki M., Migdalecki A.,	
RA Gromadka R., Noel B., Blanc I., Dessen P., Wincker P., Keller A.M.,	
RA Cohen J., Meyer E., Sperling L.;	
RT "High Coding Density on the Largest Paramacium tetraurelia Somatic	
RT Chromosome.";	
RL Curr. Biol. 14:1397-1404(2004).	
RN [2]	
RP NUCLEOTIDE SEQUENCE.	
RC STRAIN=Stock d4-2;	
RA Nowak J.K., Migdalecki A., Gromadka R., Zagulska M.;	
RT "Paramacium megabase sequencing project.";	
RT Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; CR548612; CAH03203.1; -; Genomic_DNA.	
KW Hypothetical protein.	
SQ SEQUENCE 500 AA; 56364 MW; DB47F90C86E579F0 CRC64;	
Query Match	13.0%; Score 110; DB 2; Length 500;
Best Local Similarity	22.4%; Pred. No. 14;
Matches 43; Conservative 34; Mismatches 75; Indels 40; Gaps 6	

Query Match 13.0%; Score 110; DB 2; Length 500;  
Best Local Similarity 22.4%; Pred. No. 14;  
Matches 43; Conservative 34; Mismatches 75; Indels 40; Gaps 6;

```
QY 4 KEFILLKOTGEVSELKPHRVTTVITQNGKEMSS-----TIVSEEDFILPVYKG 50
Db 11 KQDLUKKAKKEIRE-EFFQTVPIQNAESEKMKKKVPWNEHWATSDDKILIDATLG 69
QY 51 ELEKGYQFQWBIISGFEKGKAGYVIN-----LSKDTFKIPVFK-----KIEBK 95
Db 70 NEEAQREVQIYQITPANQVRSYALTQEGVPVQRPDDFEVEMFKSQMDKINLKIEKR 129
QY 96 EENKPTFPVSK---KQNPQVNSQLNESHK-----EDLQREHSQKSDSTKDV 144
Db 130 DEKNKEEEMKHKSKLKNQMKIKMRBEHKEKQNKVAIQWKKIEKSGSKARDLD 189
QY 145 ATVLKNNISSK 156
Db 190 EIIKQNNQISK 201

RESULT 18
ID Q815F3_PLAF7 PRELIMINARY; PRT; 829 AA.
AC Q815F3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PF11275c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Carlton J.M., Pain A., Fung E., White O., Barriman M., Hyman R.W.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteu M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Carlton J.M., Pain A., Fung E., White O., Barriman M., Hyman R.W.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteu M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Barriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
```

```
Q50VJ0_ENTHI PRELIMINARY; PRT; 384 AA.
Q50VJ0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE HMG box protein.
GN ORFNames=188.t00012;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Almark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAPB01000585; EAL45607.1; -; Genomic DNA.
SQ SEQUENCE 384 AA; 45464 MW; 20778965D72B019 CRC64;

Query Match 12.9%; Score 109.5; DB 2; Length 384;
Best Local Similarity 27.1%; Pred. No. 11;
Matches 39; Conservative 28; Mismatches 64; Indels 13; Gaps 4;

QY 17 ELKEHRTVTITQNGKEMSSITVSEEDFILPVYKGELEKGYQFD--GWEISFEGKKGAGY 74
Db 226 EEKTKYVEIKEDDEKTKYVEIKED-----EKKEKKSKKEDKKKEKMKNEKESDK 280
QY 75 VINLSKDTFKIPVFKIEEKEEENKPTFDVSKKQNPQVNSQLNESHKRDQLREHS 134
Db 281 KEDTKKKKKVKKSEKKDEIKKDEKKH----EKKEKTEKKPKPSEKESKKEKKS 336
QY 135 QKSDSTKD--VTATVLKNNISSK 156
Db 337 KKEDKKDEKSKKVKEDKSKKQK 360

RESULT 20
Q812K8_PLAF7 PRELIMINARY; PRT; 609 AA.
ID Q812K8_PLAF7 PRELIMINARY; PRT; 609 AA.
AC Q812K8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Large cyclophilin-like protein.
GN Name=PF11490c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Barriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
```



```
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein MAJ3P4.20
GN Name=MAJ3P4.20; Synonym=PF00465c;
OS Plasmodium falciparum (Isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jaseal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RA "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
DR EMBL; AL008970; CAA15610.2; -; Genomic_DNA.
DR PIR; T18467; T18467.
DR GO; GO:0016829; F-lyase activity; IEA.
DR GO; GO:0006397; P:RNA processing; IEA.
DR InterPro; IPR002483; PWI.
DR Pfam; PF01480; PWI; 1.
DR SMART; SM00311; PWI; 1.
KW Hypothetical protein; Lyase.
SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;
```

```
Query Match 12.8%; Score 108.5; DB 2; Length 600;
Best Local Similarity 29.3%; Pred. No. 21;
Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;
```

```
QY 37 IVSEDFILPVY-----KGELEKGYQFDGWEISGEGKK----DAGVNLNLSKDTFIKPV 87
DB 60 ILGFEDIIYECISQLKQSKKK---DGEEDKYLNAKLTNLGTGIGNKKSDFIBEL 116
QY 88 FKKI--EEKKEE-----ENKPTFDVSK-KKDNPOVHNSQLNE-----SHRK 125
DB 117 LELLNEEKKEHIADTLNENK-TNDIKVKVNEINENNVNENKDNISNKKDKEHVSHQN 175
QY 126 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISKSTTN 160
DB 176 EHNINNVNLKKEKEYTDIQDKRKHKRLSKQSDSYKKRPFNKRKTSIER-SLSNRYDE 234
QY 161 NPNK 164
DB 235 KTNK 238
```

```
RESULT 24
ID Q512T7_ENTHI PRELIMINARY; PRT; 1069 AA.
AC Q512T7_
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Fimbriae-associated protein, putative.
GN ORFNames=89.t00007;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sacheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFB01000328; EAL47849.1; -; Genomic DNA.
SQ SEQUENCE 1069 AA; 120249 MW; 1D38E4F0A8759CB4 CRC64;
```

```
Query Match 12.8%; Score 108.5; DB 2; Length 1069;
Best Local Similarity 25.5%; Pred. No. 39;
Matches 42; Conservative 34; Mismatches 60; Indels 29; Gaps 7;
```

```
QY 14 EVSELKPRVTVTTQNG-KENSSIVSEEDFILPVYKGELEKGYQFDGWEISGEGKKDA 72
DB 138 EVSKDNVNSSSTLTNGEKKLSTSLCNEQD-----ELQKSSSSSTD--NKNDRKDE 187
QY 73 GVINLSKDTFIKPVFKIEKKEENK-----PTFDVSKKKDNPO-----VNHSQLN 120
DB 188 IHFDVPLPKNEEKISMEIESKTEEEKSNLQIFSLNLSGCKDNESVEIAKVLKSNSS 247
QY 121 ESHRKEDLQREHS-QKSDSTKD-----VTATVLDKNNISKSTT 159
DB 248 NNSGEDKQDEVSCEKFDSEKKEEMIKAEVSNKEVKDKSTT 292
```

```
RESULT 25
ID Q59PE2_CANAL PRELIMINARY; PRT; 467 AA.
AC Q59PE2_
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN ORFNames=Cao19.6351;
OS Candida albicans SC5314.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=237561;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SC5314;
RX PubMed=15123810; DOI=10.1073/pnas.0401648101;
RA Jones T., Federspiel N.A., Chibana H., Dungan J., Kalman S.,
RA Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,
RA Davis R.W., Scherer S.;
RT "The diploid genome sequence of Candida albicans.";
```

RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334 (2004).

RN [2] NUCLEOTIDE SEQUENCE.  
RC STRAIN=SC5314;  
RA Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,  
RA Roberts J., Persson K., Donnelly S., Favoreto S., Tzung K.-W.,  
RA Jones T., Scherer S., Agabian N.;  
RA "Annotation of the Genome of Candida albicans";  
RL Submitted (APR-2004) to the EMBL/GenBank/DBSJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AACQ01000192; EAK92345.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 467 AA; 52829 MW; 2F4D37A2127A7253 CRC64;

Query Match 12.7%; Score 108; DB 2; Length 467;  
Best Local Similarity 20.8%; Pred. No. 17;  
Matches 50; Conservative 36; Mismatches 58; Indels 96; Gaps 9;

QY 10 KDTGEVSEL-KPHRVTVT-----IQNGKMSSTIVSEEDF----- 43  
Db 189 KSTPKTSPLRKPKPTVTPVRKMAKRPSPSATNTPEIKPKSSSPPIISSEDFLEMD 248  
QY 44 ----ILPVYKGELEKGYQDGEWISGFEKGDAGVYVNLKDTFKIPVKPKIEKKKEEN 99  
Db 249 KSTERVPIIE-----FNFDNYD--NDEKKEE--VVKSKNE--NQNTKGMESKPKPK 297  
QY 100 KPTFDVSKKNDPQVN-----HSQLNESHKHEKDLQREE 132  
Db 298 KPSSETSKTKVQPKQKQKPLSEETVLTDLDDDFKDLQLELLEEEERQPKQ 357  
QY 133 HSQKSDSKDV-----TATVLDKNNISSKSTNNPNK 164  
Db 358 QIKETKSNQSIQKKPSPTIEVDPIAFNDSDESDFDFTGKIDEGNNSSSNNNK 417

RESULT 26

Q59PL2 CANAL  
ID Q59PL2 CANAL PRELIMINARY; PRT; 467 AA.  
AC Q59PL2  
DT 10-MAY-2005 (TREMBlrel. 30, Created)  
DT 10-MAY-2005 (TREMBlrel. 30, Last sequence update)  
DT 10-MAY-2005 (TREMBlrel. 30, Last annotation update)  
DE Hypothetical protein.  
GN ORFNames=Ca019.13708;  
OS Candida albicans SC5314.  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=237561;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=SC5314;  
RX PubMed=1523810; DOI=10.1073/pnas.0401648101;  
RA Jones T., Federapfel N.A., Chibana H., Dungan J., Kalman S.,  
RA Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,  
RA Davis R.W., Scherer S.;  
RT "The diploid genome sequence of Candida albicans";  
RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334 (2004).  
RN [2]

NUCLEOTIDE SEQUENCE.

RC STRAIN=SC5314;  
RA Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,  
RA Roberts J., Persson K., Donnelly S., Favoreto S., Tzung K.-W.,  
RA Jones T., Scherer S., Agabian N.;  
RA "Annotation of the Genome of Candida albicans";  
RL Submitted (APR-2004) to the EMBL/GenBank/DBSJ databases.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AACQ01000189; EAK92416.1; -; Genomic\_DNA.  
KW Hypothetical protein.  
SQ SEQUENCE 467 AA; 52769 MW; CFBEE561B6BD8B588 CRC64;

Query Match 12.7%; Score 108; DB 2; Length 467;  
Best Local Similarity 20.8%; Pred. No. 17;  
Matches 50; Conservative 36; Mismatches 58; Indels 96; Gaps 9;

QY 10 KDTGEVSEL-KPHRVTVT-----IQNGKMSSTIVSEEDF----- 43  
Db 189 KSTPKTSPLRKPKPTVTPVRKMAKRPSPSATNTPEIKPKSSSPPIISSEDFLEMD 248  
QY 44 ----ILPVYKGELEKGYQDGEWISGFEKGDAGVYVNLKDTFKIPVKPKIEKKKEEN 99  
Db 249 KSTERVPIIE-----FNFDNYD--NDEKKEE--VVKSKNE--NQNTKGMESKPKPK 297  
QY 100 KPTFDVSKKNDPQVN-----HSQLNESHKHEKDLQREE 132  
Db 298 KPSSEASKTKVQPKQKQKPLSEETVLTDLDDDFKDLQLELLEEEERQPKQ 357  
QY 133 HSQKSDSKDV-----TATVLDKNNISSKSTNNPNK 164  
Db 358 QIKETKSNQSIQKKPSPTIEVDPIAFNDSDESDFDFTGKIDEGNNSSSNNNK 417

RESULT 27

Q4YMU4 PLABE  
ID Q4YMU4 PLABE PRELIMINARY; PRT; 662 AA.  
AC Q4YMU4;  
DT 13-SEP-2005 (TREMBlrel. 31, Created)  
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)  
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)  
DE RNA binding protein, putative.  
GN ORFNames=PB001104.03.0;  
OS Plasmodium berghei.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=5821;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolj T.W.A.,  
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,  
RA James K., Rutherford K., Harris B., Harris D., Church C.,  
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,  
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,  
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;  
RT "A comprehensive survey of the Plasmodium life cycle by genomic,  
RT transcriptomic, and proteomic analyses";  
RL Science 307:82-86 (2005).  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC EMBL; CAJ101003467; CAJ100666.1; -; Genomic\_DNA.  
DR InterPro; IPR012972; NLE.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF08154; NLE; 1.  
DR Pfam; PF00400; WD40; 8.  
DR PRINTS; PR00320; GPROTEINRPT.  
DR PRODOM; PD000018; WD40; 4.  
DR SMART; SM00320; WD40; 8.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 2.  
DR PROSITE; PS00082; WD\_REPEATS\_2; 6.  
DR PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
KW Repeat; WD repeat.  
SQ SEQUENCE 662 AA; 75536 MW; E7C8543AF5E59124 CRC64;

Query Match 12.7%; Score 108; DB 2; Length 662;  
Best Local Similarity 25.6%; Pred. No. 25;  
Matches 43; Conservative 28; Mismatches 55; Indels 42; Gaps 8;

QY 10 KDTGEVSELKPHRVTVTITQNGKMSSTIVS-----EDFILPVYKGELEKGYQD 61  
Db 472 KTKLLGHQKPVHTQFSPNGKFTASSFDKSIKRWISGDGAYLAVFRGHVGPAYKI-AW 530  
QY 62 EISGFEGKGDAGYVNLKDTFKI-----PVFKKIE-----KKEENK-PTF 103  
Db 531 SI-----DNNYIISQSDTLKLRINHLVPLLLKKKEENDEQTKNEQSEQNEHKN 583

Qy 104 DVSKK---KONPQNHSQLNESHKEDLQREHSQKSDSTKDVTATVL 148  
| : |||| | : : | : | : | : | : | :  
Db 584 DYAKTKTSKONDHANNOEDGEEKK-----KKKEKNDKIKSKIKTLL 626

RESULT 28  
ID Q9FJK9 ARATH PRELIMINARY; PRT; 470 AA.  
AC Q9FJK9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Gb|AAZ20218.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=99087489; PubMed=9872454;  
RA Nakamura Y., Sato S., Asamizu E., Kaneko H., Miyajima N.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
RT physically assigned F1 and TAC clones.";  
RL DNA Res. 5:297-308(1998).  
DR EMBL; AB015468; BAB10694.1; -; Genomic DNA.  
SQ SEQUENCE 470 AA; 53758 MW; 6D686CE7E2B35AC54 CRC64;

Query Match 12.7%; Score 107.5; DB 2; Length 470;  
Best Local Similarity 20.1%; Pred. No. 19;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;  
  
Qy 19 KPHRVTVTIOHGKMSSTIVSEEDFILPVYKGEKGVPDGWISGF-----GKK 70  
| : |||| | : : | : | : | : | : | :  
Db 82 RENRVTVDVQNSGESK-----VVQDLARRIYDE-EATSQSQRIDHPNQK 129  
| : |||| | : : | : | : | : | : | :  
  
Qy 71 DAGVINLSKDTFKIPVKPIEKKEBKENTPTDPVSKKKCN----- 111  
| : || : : : || : | : | : | : | :  
Db 130 NVGITERAFENSPIEETSHRVDDNKRRINQNFNTAAKSSENAVSRSFGADHKRAEVNGK 189  
| : || : : : || : | : | : | : | :  
  
Qy 112 PQVNHSQLNE-----SHRKEDLQREHSQKSDSTKDVTATVLQNNISSKTTNNPK 164  
| : | : : ||||| : : | : | : | : | : | :  
Db 190 PMENRDQVROTESAEKSHRKENVTKSEKPRDQGVYKTEAKDKDNKEKKEKTESINK 248  
| : | : : ||||| : : | : | : | : | : | :

RESULT 29  
ID Q64OL5\_MOUSE PRELIMINARY; PRT; 1455 AA.  
AC Q64OL5;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Sarcoma antigen NY-SAR-41  
GN Name=Ccdc118; Synonyms=4932411G06rik;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;  
OC Muridea; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN=C57BL/6; TISSUE=Head;  
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins B.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Bustow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T.I., Max S.I., Wang J., Hsieh F.,  
RA Datschenko L., Marsden K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uscin T.B., Toshivuki S., Carninci P., Prange C.,



```

QY 28 QNGKMSSTVSEEDFILPVYKGL--EKG-----YQFDGWISGPEGKDGAGYVNL-- 78
Db 19 QSGHPLKFSFWEKTHLQIYNRLREBGNKDNVEQMSPSISGTEGKKEIQMISHLQL 78
QY 79 -----SKDFIKPVFKIEKEEENKPTFDVSKKKDNPVNHSQLNESHKEDLQREH 133
Db 79 QSGKHQDSVLSANDSNLKGANEASGNGKSDENVKKSD--EENAKKSD-----EEN 134
QY 134 SOKSDSTKDVATVLDKNKISSK---TTNN 161
Db 135 KDANSNTKDAESAEGEENPVSQENQMKTLNN 165

RESULT 31
QSW5T1 TETPY
ID Q5W5T1_TETPY PRELIMINARY; PRT; 1015 AA.
AC Q5W5T1_
DT 01-FEB-2005 (TREMBLrel. 29, Created)
DT 01-FEB-2005 (TREMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TREMBLrel. 29, Last annotation update)
DE Putative P-glycoprotein (Fragment).
GN Name=mdrl;
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymenidae; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GL;
RA Camares O., Denizau F., Bamad M.;
RT "Characterisation of MDR sequence homologue in Tetrahymena
pyriformis.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AJ514918; CDS5936.2; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005524; P:ATP binding; IEA.
DR GO; GO:0016887; F:ATPase activity; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:000166; P:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR011527; ABC membrane 1.
DR InterPro; IPR001140; ABC TM transp.
DR InterPro; IPR003439; ABC transp_like.
DR Pfam; PF00664; ABC membrane; 1.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS0929; ABC TM1F; 1.
DR PROSITE; PS00211; ABC TRANSPORTER 1; 2.
DR PROSITE; PS00893; ABC TRANSPORTER_2; 2.
KW ATP-binding; Nucleotide-binding.
FT NON_TER 1
SQ SEQUENCE 1015 AA; 114219 MW; CB2E9AB73768A778 CRC64;

Query Match 12.6%; Score 106.5; DB 2; Length 1015;
Best Local Similarity 27.9%; Pred. No. 51;
Matches 43; Conservative 26; Mismatches 62; Indels 23; Gaps 6;

QY 12 TGEVSELPHRVTVTIQNGKMSSTVSEEDFILPVYKGELEKGYQFDGWISGPEGKD 71
Db 308 TGRITVIAHRLT-TIRNAD-----ILVIDKGLVQGTFD--QLIDARGKE 353
QY 72 AGYVNLKSDTFIKPVFKIEKEEENKPTFDVSKKKDNPVNHSQLNESH-RKEDLQR 130
Db 354 ALAKNQIQEKQKDEERKCEQLQEDRK--DFEERMRKSTVGKSLIMENHLTKEQIAA 410
QY 131 REHSOKSDSTKDVATVLDKNKISSKSTTNPNK 164
Db 411 EEEKEQAYFKE-----LDKNMWTFLFTWNRPER 439

```

```

RESULT 32
Q6BRW2 DEBHA
ID Q6BRW2_DEBHA PRELIMINARY; PRT; 616 AA.
AC Q6BRW2_
DT 25-OCT-2004 (TREMBLrel. 28, Created)
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Similar to CA4458|IPF8464 Candida albicans IPF8464 unknown
DE function.
GN OrderedLocusNames=DEHA0D146749;
OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36239 / CBS 767;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrans P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boisarane A., Boyer J., Cattolico L., Confanieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantreay F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikoleki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
DR EMBL; CR382136; CAG87226.1; -; Genomic DNA.
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR000749; ATP-gua_Ptrans.
KW Complete proteome.
SQ SEQUENCE 616 AA; 72143 MW; 884009B2B8B6C3CF CRC64;

Query Match 12.5%; Score 106; DB 2; Length 616;
Best Local Similarity 30.7%; Pred. No. 32;
Matches 43; Conservative 20; Mismatches 49; Indels 28; Gaps 9;

QY 46 PVYGELEKGYQFDGWISGPEGKKGAGYVNLKSDT-FIKPV----PKIEEKKKEENK 100
Db 87 PVLGRRAKPKIVLTDTIT-----KDIN-DINFSDSSEHKPIETSKTKCKTKTKTK 140
QY 101 PTFDVSKKK-----DNPO--VNHSQALNE---SHRKEDLQREHSQK-----SDSTKDV 144
Db 141 PDLDIGKLERIVADNPDEIENHSESESEIKQRKEKQRKQDKQREKLKAKNQESNNDST 200
QY 145 ATVLDK-NNISSKSTTNPN 163
Db 201 TEQPEPLKNINEKITSNEPS 220

RESULT 33
O96229 FLAF7
ID O96229_FLAF7 PRELIMINARY; PRT; 951 AA.
AC O96229_
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein PFB0680w.
GN Name=PFB0680w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99021743; PubMed=9804551; DOI=10.1126/science.282.5391.1126;

```

```

RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perteau M.,
RA Salberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteau M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrett B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AF001410; RAC71925.2; -; Genomic_DNA.
DR PIR; B71609; B71609.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 112486 MW; AC8D889358A84F4P CRC64;

Query Match 12.5%; Score 106; DB 2; Length 951;
Best Local Similarity 24.3%; Pred. No. 52;
Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;

QY 10 KDTGEVSELKPHRVTV--VTIQKEMSSSTVSEDFILPVYKGELEKGYQPDGWEI--SGF 66
DB 128 EEKINKNSDLRQNELNLQSK-----NQDI-----NKNEKGKQ-----DISNSNA 170

QY 67 EGKQDAGVYINLSKDTFKIPVKFKIEKKE--ENKPTFD-----VSKKKNP 112
DB 171 ENKKD-----VKEGVKLEEKKEKBEKISDDHKVKEKSDDKHKVEKNSDDH 218

QY 113 QVNHSQLNESHKRELQREHSQKSDSTKVTVTLDKNNIS---SKSTTNPN 164
DB 219 KVEENKSDDKHKIEEVKVEHEEEDDEE-----DKKEKKSENKKNKDNK 262

RESULT 34
Q7RJ3C PLAYO PRELIMINARY; PRT; 605 AA.
AC Q7RJ3C;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE Hypothetical protein (Fragment).
GN Name=PY03340;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR HSSP; P16649; 1ERJ.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 4.
DR PROSITE; PS00678; WD REPEATS_1; 2.
DR PROSITE; PS00682; WD REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 674 AA; 76535 MW; E7521B469FE8E0F7 CRC64;

Query Match 12.4%; Score 105.5; DB 2; Length 674;
Best Local Similarity 23.5%; Pred. No. 39;
Matches 39; Conservative 28; Mismatches 64; Indels 35; Gaps 6;

```

```

RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABL01000954; EAA22905.1; -; Genomic_DNA.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 605 AA; 69357 MW; 74035FFP4D548483 CRC64;

Query Match 12.4%; Score 105.5; DB 2; Length 605;
Best Local Similarity 24.6%; Pred. No. 34;
Matches 43; Conservative 37; Mismatches 58; Indels 37; Gaps 8;

QY 9 NKDTGEVSELKPHRVTVT--IQNGKEMSSSTVSE--EDFILPVYKGELEKGYQPDG 60
DB 369 NQDVFDLSALKEYGDSIKGHEEYKEVYTSLSNDIQEHLIEDVWLPLY--ETYSNY---- 422

QY 61 WEISGFEGKQDAGVYINLSKDTFKIPVKFKIEKKE--ENKPTFDVSK-----KKDNP 112
DB 423 -----ADYYENFDIMEYLKVELEANEPQKTRTLEDKPPINVLFTPLPDRTP 471

QY 113 QVNHSQLNESHKRELQREHSQKSDSTKVTVTLDKNNIS---SKSTTNPN 163
DB 472 VQQTQMQDAPAKISDQGNDSQESLSQKIDSTYSGKDKISDIPSEKRSQSPD 526

RESULT 35
Q7RL87 PLAYO PRELIMINARY; PRT; 674 AA.
AC Q7RL87;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Notchless-related.
GN Name=PY02598;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR HSSP; P16649; 1ERJ.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 4.
DR PROSITE; PS00678; WD REPEATS_1; 2.
DR PROSITE; PS00682; WD REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 674 AA; 76535 MW; E7521B469FE8E0F7 CRC64;

Query Match 12.4%; Score 105.5; DB 2; Length 674;
Best Local Similarity 23.5%; Pred. No. 39;
Matches 39; Conservative 28; Mismatches 64; Indels 35; Gaps 6;

```

```
QY 10 KDTGEVSELKPHRVTTVTIONGKEMSGTIVSE-----EDFILPVYKGELEKGVQPGW 61
DB 481 KTRLLGHQKPVHRTQSPNGKFIASSSPDKSIRIWSGIDGTLYLAVFRGHVGPAYKI-AW 539
QY 62 BISGPEGKADAGYVNLKSDTPIK-----PVFKKIEE-----KKEENKPTFD 104
DB 540 SI-----DNYIISCSQDSTLKLWRINHLVFLKKKEGSDQPKSPKNDQKDKD 592
QY 105 V--SKKNQPVNHSQNLNESHKEDLQREHSQKSDSTKDVATVYL 148
DB 593 XKDDQKQDKNDNDHANNQEDGEKKKKKKEKDKKIKSKIKTL 638

RESULT 36
Q54GS1_DICDI
ID Q54GS1_DICDI PRELIMINARY; PRT; 1550 AA.
AC Q54GS1_
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=DDB0188660;
OS Dictyostelium discoideum (Slime mold). Dictyostelium.
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
ON NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tungal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero P.,
RA Bankier A.T., Lehman R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Parbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsey R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Loulseged H., Mungali K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaulsky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.,
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0:0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFI01000221; EAL62484.1; --; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1550 AA; 174527 MW; D6DSDFF547DC48E6 CRC64;

Query Match 12.4%; Score 105; DB 2; Length 1550;
Best Local Similarity 25.4%; Pred. No. 1e+02;
Matches 46; Conservative 24; Mismatches 59; Indels 52; Gaps 6;

QY 20 PHRVTTVTIONGKEMSGTIVSEDFILPVYKGELEK--GYQFGWEISGFEKGDAGYVI- 76
DB 1328 PHK-----LKKGNRGSSSNLSILPTLEDHLKRTVSPKVDMMFTSPFDDTLDLKELIS 1383
QY 77 -----NLKDTFFIKPVKKIEKKEENKPTFDVS----- 106
DB 1384 TDELANNNNNNNNNNNNPFSNKRNSNEVKVQPVQOEIEKEKEKNIGEDVEKETS 1443
QY 107 -----KKKDNQVNHVSQNLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 160
DB 1444 IQEKIKDNDNNNNNIVEKEEIEIKDNIKEIKQDSDNNKEI-----NN--DNSTTN 1494
QY 161 N 161
```

```
DB 1495 N 1495
RESULT 37
Q9NFV9_PLAFA
ID Q9NFV9_PLAFA PRELIMINARY; PRT; 329 AA.
AC Q9NFV9_
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN Name=map3;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=7G8;
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
RT Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252287; CAB85901.1; --; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
FT NON TER 1
FT NON_TER 329
SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;

Query Match 12.3%; Score 104.5; DB 2; Length 329;
Best Local Similarity 24.3%; Pred. No. 21;
Matches 43; Conservative 30; Mismatches 63; Indels 41; Gaps 8;

QY 1 TTVEFILN-KDTGEVSELKPHRVTTVTIONGKEMSGTIVSEDFILPVYKGELEKGYQFD 59
DB 144 TKTKEAFKAYEAKNAYQKANQAVLKAKEASS-----YDYL----- 184
QY 60 GWEISGPF--EGKKDAG-----YVNLKSDTFFIKPVKKIEKKEENKPTFDVSKKDN 112
DB 185 GWEFGGVPPEHKKEENMLSHLVSSKKNISKENDVDLDE--KEEAEETEHEELEEKNE 243
QY 113 QVNHVSQNLNESHKEDLQREH-----SOKSDSTKDVATVLDKNNISKSTTN 161
DB 244 EETSEISEDEEEEEEKEEKEKQAKSQSNENNNDQKDMRA-----QNLISKNNNN 295

RESULT 38
Q4YVY2_PLABE
ID Q4YVY2_PLABE PRELIMINARY; PRT; 736 AA.
AC Q4YVY2_
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=PB000556.02.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
```

**QY** 1 TTVKPEFLNKDGTGVSELKPHRVVTITQNGKEWSSSTIVSEDFILPVVKGLEKGYQFDG 60  
||||| :  
**Db** 27 TTYYKA-VVEETKYVEEESKP-----EGVEKSASFKEBSDFPADLKESEKK----- 70  
||| :  
**QY** 61 WEISGFEGKKDAGVVNLNSKDTPFKVPFK--LEEKEENKTFTDVSKKDNPNVNH 117  
||| :  
**Db** 71 -AUSDLSKSLKEBAIVN----TLTKTKGSSPMKEKKEEVVKPEAEVEKKKE--EAABE 123  
||| :  
**QY** 118 QLNESHKRDLQREHSQKSDSTKDTATVLDKNNISS 155  
||| :  
**Db** 124 KVEEEKSEAVTTEAPKAETVEAVTTEEIIPKEEVI 161  
||| :

**RESULT 40**

**Q9SA84 ARATH**

ID Q9SA84\_ARATH PRELIMINARY; PRT; 540 AA.

AC Q9SA84\_2000 (Tremblurel. 13, Created)

DT 01-MAY-2000 (Tremblurel. 13, Last sequence update)

DT 01-FEB-2005 (Tremblurel. 29, Last annotation update)

DE TS18.14 protein (Hypothetical protein Atlg30690).

GN Name=TS18.14; Synonyms=Atlg30690;

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosoids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Vysochekskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,

RA Li J., Kremensheka I., Luross J., Ngan I., Gonzalez A., Alcafi H.,

RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,

RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,

RA Davis R.W., Ecker J.R., Pedersen N.A., Theologis A.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP NUCLEOTIDE SEQUENCE.

RA Theologis;

RA Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP NUCLEOTIDE SEQUENCE.

RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hauan V.W., Lee J.M.,

RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,

RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,

RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,

RA Ecker J.R., Theologis A.;

RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC007060; AAD25756.1; - ; Genomic\_DNA.

DR EMBL; BT000959; AAN41359.1; - ; mRNA.

DR PIR; D86432; D86432.

DR GO; GO:0005622; C:intracellular; IEA.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0008220; F:protein carrier activity; IEA.

DR GO; GO:0005215; F:transporter activity; IEA.

DR GO; GO:0006886; P:intracellular protein transport; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR001251; CRAL\_TRIO\_C.

DR InterPro; IPR008273; CRAL\_TRIO\_N.

DR InterPro; IPR000348; Emp24\_GP25L\_p24.

DR InterPro; IPR009038; GOLD.

DR InterPro; IPR001071; RetBind/tocTrans.

DR Pfam; PF00650; CRAL\_TRIO; 1.

DR Pfam; PF03765; CRAL\_TRIO\_N; 1.

DR Pfam; PF01105; EMP24\_GP25L; 1.

DR PRINTS; PR00180; CRETINALDHP.

DR SMART; SMO0516; SEC14; 1.

DR PROSITE; PS50191; CRAL\_TRIO; 1.

DR PROSITE; PS50866; GOLD; 1.

KW Hypothetical protein.

SQ SEQUENCE 540 AA; 61189 MW; 0C2590D518ACFB58 CRC64;

Query Match 12.3%; Score 104; DB 2; Length 540;  
Best Local Similarity 25.3%; Pred. No. 39;

```

Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;
Qy 1 TTWKEFILNKDTGEVSELPKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
Db ||| : : | | | | | : : | | : : | :
27 TTVKA-VVEETKVEEDEXP-----EGVEKSASFKEESDFFADLKSEKK----- 70
Qy 61 WEISGPEGKKGAGYVNLSDTPIKPVFKK---IEKKKEENKPTFDYKKKDNQVNH 117
Db : : : : | | | | | : : : | | | | | : : : | | : :
71 -ALSDLKSEEAIVDN-----TLLTKKESSPMKKEEVVYKPEAEVEKKKB--EAAEE 123
Qy 118 QLNESHKEDLQREEHQSQSDSTKQVTVATVLDKNNISS 155
Db : : : : | | : : | | : : | : : :
124 KVEEKKSEAVVTERAPKAETVEAVVTEIIPKEVTT 161

```

Search completed: April 24, 2006, 14:59:38  
Job time : 100.611 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.  
OM protein - protein search, using sw model  
Run on: April 24, 2006, 14:50:52 ; Search time 17.2438 Seconds  
(without alignments)  
915.083 Million cell updates/sec  
Title: US-10-067-385-8\_COPY\_610\_773  
Perfect score: 848  
Sequence: 1 TTVKEFILNKDTGEVSELKP.....ATVLDKNNISSKSTNNPNK 164  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : PIR 80.\*  
1: piri.\*  
2: piri2.\*  
3: piri3.\*  
4: piri4.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	848	100.0	2140 2 F95074 serine proteinase,
2	845	99.6	2144 2 A97942 metalloproteinase
3	111	13.1	1038 2 JC5497 claustrin - chicke
4	110	13.0	558 2 T18467 hypothetical prote
5	106	12.5	665 2 B71609 hypothetical prote
6	104	12.3	540 2 D86432 hypothetical prote
7	103.5	12.2	325 2 T18283 hypothetical prote
8	101.5	12.0	312 2 G81339 probable membrane
9	100	11.8	622 2 A90570 lipoprotein impor
10	99.5	11.7	385 2 T20410 hypothetical prote
11	99	11.7	211 2 T25911 hypothetical prote
12	99	11.7	219 2 B72291 hypothetical prote
13	97.5	11.5	614 2 A84152 hypothetical prote
14	97.5	11.5	1345 2 S46817 hypothetical prote
15	97	11.4	988 2 T14188 hypothetical prote
16	96.5	11.4	456 2 T05612 hypothetical prote
17	96	11.3	535 2 T37189 hypothetical prote
18	96	11.3	2500 2 G71609 hypothetical prote
19	95.5	11.3	644 2 T47835 hypothetical prote
20	95	11.2	348 2 I37271 cylicin II - human
21	95	11.2	1397 2 T10466 DNA topoisomerase
22	94.5	11.1	867 2 T27136 hypothetical prote
23	94.5	11.1	871 2 T27135 hypothetical prote
24	94.5	11.1	3724 2 T18427 hypothetical prote
25	94	11.1	210 2 T28771 hypothetical prote
26	93.5	11.0	645 2 E89883 conserved hypothet
27	93.5	11.0	649 2 S42488 dnak-type molecula
28	93	11.0	629 2 G96542 hypothetical prote
29	93	11.0	2464 1 QRMSP1 microtubule-associ

ALIGNMENTS

RESULT 1  
P95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: P95074  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Ha  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap  
neon, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morriso  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: P95074  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2140 <KUR>  
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match			
Best Local Similarity 100.0%; Score 848; DB 2; Length 2140;			
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	TTVKEFILNKDTGEVSELKPHRVTTTQNGKMSSTIVSEEDPILPVYKGLKGYQFDG	60
Db	1943	TTVKEFILNKDTGEVSELKPHRVTTTQNGKMSSTIVSEEDPILPVYKGLKGYQFDG	2002
QY	61	WEISGFEKKDAGVYINLSKDTFKVPKTEKKKEENKPTFDVSKKDNPNVHNSQLN	120
Db	2003	WEISGFEKKDAGVYINLSKDTFKVPKTEKKKEENKPTFDVSKKDNPNVHNSQLN	2062
QY	121	ESHRKEDIQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK	164
Db	2063	ESHRKEDIQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK	2106
RESULT 2			
A97942			
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)			
C:Species: Streptococcus pneumoniae			
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004			
C:Accession: A97942			
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;			
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;			
J. Bacteriol. 183, 5709-5717, 2001			
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.			
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.			
A:Reference number: A97872; MUID:21429245; PMID:11544234			

A:Accession: A97942  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KUP>  
A:Cross-references: UNIPROT:Q8DQ7; UNIPARC:UPI00000E3490; GB:AE007317; PIDN:AAK99365.1;  
C:Genetics:  
A:Gene: prta  
C:Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 845; DB 2; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 2.3e-53;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTVKPILNDTGEVSELKPHRVTTIQTNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60  
Db 1947 TTVKPILNDTGEVSELKPHRVTTIQTNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2006

Qy 61 WEISFEGKDGAGYVNLKDTFIKVPFKIEEKEEENKPTFDVSKKDNPNVNSQLN 120  
Db 2007 WEISFEGKDGAGYVNLKDTFIKVPFKIEEKEEENKPTFDVSKKDNPNVNSQLN 2066

Qy 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164  
Db 2067 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2110

RESULT 3  
JCS497  
claustrin - chicken  
N:Alternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence revision 12-Sep-1997 #text\_change 09-Jul-2004  
C:Accession: JCS497; PC4334; S37561  
R:Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JCS497; MUID:94157526; PMID:7906711  
A:Accession: JCS497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BURL>  
A:Cross-references: UNIPROT:Q90784; UNIPARC:UPI00000FC87F; EMBL:X67778; NID:G406318; PID  
A:Accession: PC4334  
A:Molecule type: protein  
A:Residues: 79-83;299-412;485-502 <BUR2>  
A:Cross-references: UNIPARC:UPI000017BPF3; UNIPARC:UPI000017BPF4; UNIPARC:UPI000017BPF5  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervo  
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

Query Match 13.1%; Score 111; DB 2; Length 1038;  
Best Local Similarity 23.7%; Pred. No. 2;  
Matches 44; Conservative 28; Mismatches 58; Indels 56; Gaps 5;

Qy 3 VKPEFILNDTGEVSELKPHRVTTIQTNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISFEGKDGAGYVNLKDTFI 83  
Db 528 VQQAALKORTDSKESLPPAAKTTTKDQCKQRNLKKKMSLSQLVQOLEKPKQLESKEKTP 587

Qy 30 -----GKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISFEGKDGAGYVNLKDTFI 83  
Db 588 VKKERAVKETIIVAKDV-----TTKEQIGKSETSEKQASEKQDVKPKVTKES 639

Qy 84 IKPVFK-KTEEKKEEENKPTFDVSKKDNPNVNSQLNESHKEDLQRE-----EHSQKS 137  
Db 640 VKKEVKAKPEEKDEKEKPKKEVSKKEEKLPI---KGEKPKKEDIKKEVKKVEKKEK 696

Qy 138 DSTKDV 143  
Db 697 EAKKEV 702

RESULT 4  
TI8467  
hypothetical protein C0465C - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: TI8467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: TI8467  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: UNIPROT:O77355; UNIPARC:UPI000017CC2E; EMBL:AL008970; NID:e1407852;  
C:Genetics:  
A:Map position: 3  
A:Introns: 84/1; 160/1  
A:Note: C0465C

Query Match 13.0%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 1.2;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

Qy 37 IVSEEDFILPVY-----KGELEKGYQFDGWEISFEGK-----DAGYVNLKDTFIKPV 87  
Db 60 ILGFEDDILYEYCISQLKQSKK--KADGEBDKYLNKAKLKINLTGPIGNKKSDFIEEL 117

Qy 88 PKKI--BEKKEE-----ENKPTFDVSK-KKONPQVNSQLN-----SHRK 125  
Db 118 LELLINBEKKEHTADTLNENK-TNCJIKVKVNEINENYVYNNKQISNKDKHVSQHN 176

Qy 126 E-----DLQREH-----SQKSDTK-----DVTATVLDKNNISSKSTTN 160  
Db 177 EHNINNVNLKKEKEYTDIQDRKHRSLSQKSDSYKRPFPNPKRTKSTIER-SLSNRYDE 235

Qy 161 NPNK 164  
Db 236 KTNK 239

RESULT 5  
B71609  
hypothetical protein PF0680w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-665 <GAR>  
A:Cross-references: UNIPROT:O96229; UNIPARC:UPI000017B60A; GB:AE001410; NID:AE001362; NID  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PF0680w

Query Match 12.5%; Score 106; DB 2; Length 665;  
Best Local Similarity 24.3%; Pred. No. 2.8;  
Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;

Qy 10 KDTGEVSELKPHRV-TIQTNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGF 66  
Db 127 EEKNKINKSLHRQNELNLQSGK-----NEQDI-----NKNKKGKQ-----DISNSNA 169

Qy 67 EGKDGAGYVNLKDTFIKVPFKIEEKE-----EENKPTD-----VSKKDNPN 112  
Db 170 ENKED-----VKGEVKELEKKEEKEKTSDDHKVEENKKSDDHKVEENKKSDDH 217

Qy 113 QVNSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164



A;Status: Preliminary  
A;Molecule type: DNA  
A;Residues: 1-62 <RUR>  
A;Cross-references: UNIPROT:Q980A1, UNIPARC:UPI00000C80A1, GB:AL445566, PID:91  
A;Experimental source: strain UAB CTIP  
C;Genetics:  
A;Gene: MYPU\_4650  
A;Genetic code: SGC3



Best Local Similarity 27.3%; Pred. No. 10;  
Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;  
QY 16 SELKPHRVTTIQNGKMSSTIVSE-----BDFILPVYKGELEKGYQFDGW-----EISG 65  
DB 386 SBPKPE--TYTLQTAIQM--TPIVNEVSPQTRBEFL-----ARAHQLDGWADVSKVTY 435  
QY 66 FEGKDGAGVIVNLSKDTFKPVFKIEKK--BEENKTPFDVSKK-----DNPQVNHSL 119  
DB 436 FAGRTASQL-----GKIEEKLQDKYNNYTFDFKKEVVNVQSPIKSTGAL 482  
QY 120 NESHRKEDLQREHSQKSDTKDVTATVLD 149  
DB 483 GKVIGIVDLGRVHKHGEDLERLSKSQIE 512

## RESULT 14

S46817  
hypothetical protein YHR080c - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Oct-2004  
C:Accession: S46817  
R:Pavello, T.  
submitted to the EMBL Data Library, June 1994  
A:Description: The sequence of S. cerevisiae cosmid 9205.  
A:Reference number: S46795  
A:Accession: S46817  
A:Molecule type: DNA  
A:Residues: 1-1345 <Fav>  
A:Cross-references: UNIPROT:P38800; UNIPARC:UPI000013B2B1; EMBL:U10556; NID:G500825; PID  
C:Genetics:  
A:Cross-references: SGD:S0001122  
A:Map position: 8R  
C:Superfamily: uncharacterized conserved protein  
C:Keywords: transmembrane protein

Query Match 11.5%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 25;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;  
QY 49 KGELEKGYQFDGWISGPEGKK-DAGYVIVNLSKDTFKPVFKIEKKKEENKPTFDVSK 107  
DB 1109 KGAIEKG-----SVGEQKVSDYMLSELRDII-----SRASKKPVKKVMK 1149  
QY 108 KCONPQVNHSLNESHRKEDLQREHSQKSDTKDVTATVLD--KNNISKSSTTN---NP 162  
DB 1150 SHDKHPPHFKVE-----QKSSESRKSDNDKDLTHILDVQNNFSSSEIFNMKLLSP 1201  
QY 163 NK 164  
DB 1202 QK 1203

## RESULT 15

T14188  
hypothetical protein T28D5.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T14188  
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: Z17931  
A:Accession: T14188  
A:Molecule type: DNA  
A:Residues: 1-988 <BEV>  
A:Cross-references: UNIPROT:Q9STN4; UNIPARC:UPI00000A3618; EMBL:AL109819  
A:Experimental source: cultivar Columbia; BAC clone T28D5  
C:Genetics:  
A:Gene: ATSP:T28D5.30  
A:Map position: 4  
A:Introns: 162/3; 201/3; 416/3; 438/3; 460/3; 482/3; 504/3; 519/3; 534/3; 559/3; 579/3;  
C:Superfamily: Arabidopsis thaliana hypothetical protein T21C14.40

Query Match 11.4%; Score 97; DB 2; Length 988;  
Best Local Similarity 22.1%; Pred. No. 19;  
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;  
QY 29 NGKMSSTIVSEEDFILPVYKGELEKGY-----OPDGWISGFEKGDAGYVIVNLSKDTF 83  
DB 393 NGRQNSNVQSVDEILSYTTDKVPVSGVGLNVSRDIVELVEDDVRSGAGLSPNVQRDN- 441  
QY 84 IKPVFKIIEKKK-----BEENKTPFDVSKKONPQVNHSLNESHRKEDLQREE 132  
DB 442 VEPVGDDVRSSGDSPNPASANNVREGPATFDIMESDNPGRDNVAPMEDHIRSEVQLSP 501  
QY 133 HSQKSDSTKDTV--ATVLDKNNISKSSTTNPNK 164  
DB 502 HVL---GAKDVTDSVSDPTDKRVGNDVTDASDPT 532

## RESULT 16

T05612  
hypothetical protein F9D16.270 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T05612  
R:Bevan, M.; Wedler, H.; Wedler, B.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05612  
A:Molecule type: DNA  
A:Residues: 1-456 <BEV>  
A:Cross-references: UNIPROT:Q9SUP7; UNIPARC:UPI00000A81P5; EMBL:AL035394  
A:Experimental source: cultivar Columbia; BAC clone F9D16  
C:Genetics:  
A:Map position: 4  
A:Introns: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
A:Note: F9D16.270

Query Match 11.4%; Score 96.5; DB 2; Length 456;  
Best Local Similarity 25.0%; Pred. No. 8.8;  
Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;  
QY 10 KDTGEVSELKP-HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGPEG 68  
DB 245 KKKKKDPLKPKHPVSAFLVYVANERRAALREKNSVVEVAK-----ITGEWKNLSD 296  
QY 69 KKDAGY--VINLSKDTFKPVFKIEKKKEENKPTFDVSKKONPQVNHSLNESHRKE 126  
DB 297 KKKAPYKVKAKNKETVYQ-AMEBYKRTKEE-----ALSQKE-----BEELKLHKQE 345  
QY 127 DLQREHSQKSDTKDVTATVLDKNNISKSSTTN--PNK 164  
DB 346 ALQMLKKKEKTDN-----LIKKEKATKKKKNNVDPNK 378

## RESULT 17

T37189  
hypothetical protein C02H7.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37189  
R:Leimbac, D.; Mixx, M.  
submitted to the EMBL Data Library, February 1996  
A:Description: The sequence of C. elegans cosmid C02H7.  
A:Reference number: Z20523  
A:Accession: T37189  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-535 <LEI>  
A:Cross-references: UNIPROT:Q17595; UNIPARC:UPI000008019E; EMBL:U49945; PIDN:AAC47924.1  
A:Experimental source: strain Bristol N2; clone C02H7  
C:Genetics:  
A:Gene: CESP:C02H7.1  
A:Map position: X  
A:Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 11.3%; Score 96; DB 2; Length 535;  
Best Local Similarity 21.3%; Pred. No. 11;  
Matches 35; Conservative 28; Mismatches 61; Indels 40; Gaps 5;  
QY 11 DTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDGEWISGFEKKG 70  
DB 77 DGSLKNVKAII---SGDAET-----NKMQLMGTNATSFNSRN 116  
QY 71 DAGYVNLKOTFIKPVFKIEKKEENKPTFDVSKKNDPQNVHSQLNESHKEDLQR 130  
DB 117 GTG-----EKKKKKKVKEDKGDDEKST---TKRSSKKTHERKEKSEKKKSAAE 166  
QY 131 EHSQKSDSTK-----DVTATVLDKNNISSKSTTNNPNK 164  
DB 167 KEKKKKSSSKERHKSSDRSSSEKSKSEKKEKSTTDEKPK 210  
RESULT 18  
G71609  
hypothetical protein PF0650w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C>Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: G71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: G71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2500 <GAP>  
A:Cross-references: UNIPROT:O96223; UNIPARC:UPI0000078301; GB:AE001408; GB:AE001362; NID  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PF0650w  
Query Match 11.3%; Score 96; DB 2; Length 2500;  
Best Local Similarity 26.1%; Pred. No. 64;  
Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;  
QY 19 KPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDGEWISGFEKKGAGVNL 78  
DB 2173 KPIKIT---ENKK-----NEGNEILKKYSIENEKKNYDKENQNCILDKDTQCNVT 2223  
QY 79 -----SKDTFIKPVFKIEKKEENKPTFDVSKKNDPQNVHSQLNE-----SHRKED 127  
DB 2224 KEKNLNDKKSPFNIKVKLEBEKSDDKRD---DKKNDTREKNLNDKKSPFNIKVK 2280  
QY 128 LQREHSQKSDSTKQVATVLDKNNISSKSTTNNPNK 164  
DB 2281 LEEEE---KSDDKRD-----DKKNDTREKNLNDK 2308  
RESULT 19  
T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C:Accession: T47835  
R:Nyakatura, G.; Partmann, B.; Dauner, D.; Starr, W.; Holland, R.; Weicheelgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24475  
A:Accession: T47835  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <NYA>  
A:Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI000009E0AD; EMBL:AL138658  
A:Experimental source: cultivar Columbia; BAC clone T209  
C:Genetics:  
A:Map position: 3  
A:Introns: 158/2; 329/3

A;Note: T209.90

Query Match 11.3%; Score 95.5; DB 2; Length 644;  
Best Local Similarity 23.4%; Pred. No. 15;  
Matches 34; Conservative 22; Mismatches 58; Indels 31; Gaps 4;  
QY 28 QNGKEMSTIVSEB---DFILPVYKGELEKGYQDGEWISGFEKKGADGVNLNLSKDTF 83  
DB 519 ENSKTEKTVADKKKSVADFLKRIKKNSPQKG-----ETTSKNQKKNDGNV----- 565  
QY 84 IKPVFKKIEKKEENKPTFDVSKKNDPQNVHSQLNESHKEDLQREHSQKSDS----- 139  
DB 566 -----KENDHOKKSDGNVKKENSKVKPRELSTGKKKVEYNNNNSSSKRKQ 615  
QY 140 TKDVTATVLDKNNISSKSTTNNPNK 164  
DB 616 TKETAEVATGREGSKDKDKQPK 640

RESULT 20

I37271  
Cylcin II - human  
C:Species: Homo sapiens (man)  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 09-Jul-2004  
C:Accession: I37271; S52774  
R:Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.  
Exp. Cell Res. 218, 174-182, 1995  
A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the  
A:Reference number: I37271; MUID:95255491; PMID:7737358  
A:Accession: I37271  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-348 <HES>  
A:Cross-references: UNIPROT:Q14093; UNIPARC:UPI0000128C36; EMBL:246788; NID:g758586; PII

Query Match 11.2%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 8.4;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;  
QY 49 KGELEKGYQDGEWISGFEKKGADGVNLNLSKDTFIKPVFKKIEKKEENKPTF---DV 105  
DB 205 ESEGEK---GTSEKSKKKKDS---KKGKSAIBLQAVKADEKDEKDKANKGDE 256  
QY 106 SK--KNDPQNVHSQLN-----ESHKEDLQREHSQKSDSTKD---VTATVLDKNNI 153  
DB 257 SKDAKKAKEIKKGGKKKPSSTSDSKDVKKE---SKDATKDAKVAKKDKTEKESA 313  
QY 154 SSK 156  
DB 314 DSK 316

RESULT 21

T10466  
DNA topoisomerase (ATP-hydrolyzing) (SC 5.99.1.3) II - malaria parasite (Plasmodium falci  
C:Species: Plasmodium falciparum  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C:Accession: T10466  
R:Cheesman, S.J.  
submitted to the EMBL Data Library, September 1995  
A:Reference number: Z17031  
A:Accession: T10466  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1397 <CHE>  
A:Cross-references: UNIPARC:UPI00000006D7; EMBL:X79345; NID:g994807  
C:Genetics:  
A:Gene: TopoII  
A:Map position: 14  
A:Superfamily: eukaryotic type II DNA topoisomerase; phase T4 DNA topoisomerase (ATP-hyd  
C:Keywords: ATP; DNA binding; isomerase; nucleus  
Query Match 11.2%; Score 95; DB 2; Length 1397;

T27135  
hypothetical protein Y53C12B.3b - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T27135  
R:Kershaw, J.; Lennard, N.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20316  
A:Accession: T27135  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-871 <Wt>  
A:Cross-references: UNIPROT:O62486; UNIPARC:UPT00000079598; EMBL:Z99278; PIDN:CAB16492.1;  
A:Experimental source: clone Y53C12B  
C:Genetics:  
A:Gene: CRSP:Y53C12B.3b  
A:Map position: 2  
A:Introns: 100/3; 177/3; 218/1; 423/3; 717/2; 867/3

T28771  
hypothetical protein E03H12.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 23-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T28771  
R:Nelson, J.; Wohldmann, P.; Sansone, J.  
submitted to the EMBL Data Library, June 1997  
A:Description: The sequence of C. elegans cosmid E03H12.  
A:Reference number: Z20520  
A:Accession: T28771  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-210 <NLE>  
A:Cross-references: UNIPROT:O02124; UNIPARC:UPI000007DF6; EMBL:AF0000299; P  
A:Experimental source: strain Bristol N2; clone E03H12  
C:Genetics:  
A:Gene: CESP:E03H12.5  
A:Map position: 4  
A:Introns: 30/2; 201/3









Best Local Similarity 32.6%; Pred. No. 11;  
Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;

Qy 86 PVFKKIKKK---EENKPTFDVSK-----KDNPQNHSQLNESH----- 124  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 61 PVPKVEKKKEEKKADDEKKTKTEKKDKKSKTBEKDKISKVKKTQTTSERKK 120  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 125 ---KEDLQRHSHSQSDTKDVTATVLDKNNSK 156  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 121 DEKKDBKBESKEKSDEEKKCKDEVKDEK 155  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 36

TG3030  
gelsolin-related protein GRP125 - slime mold (*Dictyostelium discoideum*)  
C:Species: *Dictyostelium discoideum*  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T30330  
R/Stocker, S.; Hiery, M.; Marriott, G.  
Mol. Biol. Cell 10, 161-178, 1999  
A>Title: Phototactic migration of *Dictyostelium* cells is linked to a new type of gelsolin  
A/Reference number: Z20823; MUID:99096692; PMID:9880334  
A/Accession: T30330  
A>Status: preliminary; translated from GB/EMBL/DDBB  
A/Molecule type: DNA  
A/Residues: 1-1087 <STO>  
A/Cross-references: UNIPROT:O96923; UNIPARC:UPI000007D5CC; EMBL:U95159; NID:g4100185; PID:  
C/Genetics:  
A/Introns: 137/1

Query Match 10.7%; Score 91; DB 2; Length 1087;  
Best Local Similarity 24.4%; Pred. No. 58;  
Matches 43; Conservative 35; Mismatches 76; Indels 22; Gaps 7;

Qy 8 LNKDTGVSEVLKPRHTVTTLQNGKMSSTI-----VSREDFILPVYKGEL--EKGVQPD 59  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 721 LQKEBLEKLKQQEQEQEQKNNKI VEEVKEEVKEEDVKEEVKEEVKEEVKEE 780  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 60 GWGISGPEGKKDAGVIVNLSKDPFPIKFVFKKIBEKKEEENKPTFDVSKKD-----NP 112  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 781 VKSAKETKEEIKEYN--DEATEVKNQVEEVKEEVKEEVKEEVKEEVKEE 839  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Qy 113 QVNHSQLNESHKEDLQ--REHSQ--KSPSTKDVTATVLDKNNSKSTTNPNK 164  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :  
Db 840 EVKEEVKEEVKEEVKEEVKEEVKEEVKQDKBEEVNEEI--KEETKEEETKEDDNK 893  
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

RESULT 37

S66040  
serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4) *dacA* - *Bacillus subtilis*  
N/Alternate names: penicillin-binding protein 5  
C:Species: *Bacillus subtilis*  
C>Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004  
C/Accession: S66040; I39830; A92275; A92307; B61335; D69612; A23307  
R/Ogasawara, N.; Nakai, S.; Yoshikawa, H.  
DNA Res. 1, 1-14, 1994  
A>Title: Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromo-  
some  
A/Reference number: S65967; MUID:96051385; PMID:7584024  
A/Accession: S66040  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Molecule type: DNA  
A/Residues: 1-443 <OCA>  
A/Cross-references: UNIPROT:P08750; UNIPARC:UPI000005PDBA; EMBL:D26185; NID:g467326; PID:  
A/Note: The nucleotide sequence was submitted to the EMBL Data Library, December 1993  
R/Rodd, J.A.; Roberts, A.N.; Johnstone, K.; Piggot, P.J.; Winter, G.; Ellar, D.G.  
J. Bacteriol. 167, 257-264, 1986  
A>Title: Reduced heat resistance of mutant spores after cloning and mutagenesis of the *E. coli*  
A/Reference number: I39830; MUID:86250602; PMID:3087956  
A/Accession: I39830  
A>Status: preliminary; translated from GB/EMBL/DDBB  
A/Molecule type: DNA  
A/Residues: 48-226, 'Q', 228-443 <RES>  
A/Cross-references: UNIPARC:UPI000016E82C; GB:M13766; NID:g142816; PIDN:AAA22375.1; PID:

[illegible]

C:Species: Melanoplus sanguinipes entomopoxvirus  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T28391  
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28391  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-670 <NA>  
A:Cross-references: UNIPROT:O9YVL2; UNIPARC:UPI00000F2D3E; EMBL:AF063866; NID:g4049647;  
C:Genetics:  
A>Note: MSV230

Query Match 10.7%; Score 90.5; DB 2; Length 670;  
Best Local Similarity 22.5%; Pred. No. 37;  
Matches 36; Conservative 34; Mismatches 69; Indels 21; Gaps 5;

QY 25 VTIQNGKMSSTI-----VSEEDFILPVYKGELEKGY-----QPDGWEISGF 66  
Db 33 VNFEKKQIISILLKFNFDKTEMCGVSEKVFQLINNKASAEKSDVDSSIDESQNSDS 92  
QY 67 EGKGDAGYVINI-SKDTFIKPVFKKIE-EKKBEENKPTFDVSKKKONPQVNHSQL-NESHR 124  
Db 93 DSDSDSGVNIIDRSQNSDSKVWINKLENESQNSDSKVNIIDRSQNSDS-KVWINKLENESQ 151  
QY 125 KEDLQREHSQKSDSTKQVATVLDKNNISSKSTNNPNK 164  
Db 152 SDSKVNIIDRSQNSDSKVNIIDRSQNSDSKVNIIDRSQNSDSK 191

RESULT 39  
S41552  
probable transcription factor SPT7 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YBR0739; protein YBR081C  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-Jan-1994 #sequence\_revision 09-Sep-1994 #text\_change 05-Oct-2004  
C:Accession: S41552; S45946; S45948; S40800; S45478; S59716  
R:Gansheroff, L.; Dollard, C.; Tan, P.; Winston, F.  
submitted to the EMBL Data Library, July 1993  
A:Reference number: S41552  
A:Accession: S41552  
A:Molecule type: DNA  
A:Residues: 1-1332 <NA>  
A:Cross-references: UNIPROT:P35177; UNIPARC:UPI000005310C; EMBL:L22537; NID:g349189; PID:R.Steenma, H.Y.; van der Aart, Q.J.M.  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45932  
A:Accession: S45946  
A:Molecule type: DNA  
A:Residues: 1-1332 <STB>  
A:Cross-references: UNIPARC:UPI000005310C; EMBL:Z35950; NID:g536341; PIDN:CAA85026.1; PID:R.Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.; Urrestazu, A.; Vissers, S.  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45893  
A:Accession: S45948  
A:Molecule type: DNA  
A:Residues: 1-835 <AND>  
A:Cross-references: UNIPARC:UPI000017A10D; EMBL:Z35950; MIPS:YBR081C  
R:Haynes, S.R.; Dollard, C.; Winston, F.; Beck, S.; Trowsdale, J.; David, I.B.  
Nucleic Acids Res. 20, 2603, 1992  
A:Title: The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins.  
A:Reference number: S40800; MUID:92285152; PMID:1350857  
A:Accession: S40800  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 463-523 <HAY>  
A:Cross-references: UNIPARC:UPI0000168E4B; EMBL:M87651; NID:g172683; PIDN:AAA35087.1; PID:R.van der Aart, Q.J.M.; Barthe, C.; Doignon, F.; Aigle, M.; Crouzet, M.; Steensma, H.Y.  
Yeast 10, 959-964, 1994  
A:Title: Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyces cerevisiae chromosome V.  
A:Reference number: S45462; MUID:95076715; PMID:7985423

A:Accession: S45478  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-624, 'LRGKKRKI', 633-1332 <VAN>  
A:Cross-references: UNIPARC:UPI000017A10E; EMBL:X76294  
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1993  
R:Gansheroff, L.J.; Dollard, C.; Tan, P.; Winston, F.  
Genetics 139, 523-536, 1995  
A:Title: The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein important for chromosome segregation.  
A:Reference number: S54985; MUID:95229044; PMID:7713415  
A:Accession: S54985  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1332 <GAW>  
A:Cross-references: UNIPARC:UPI000005310C; EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:R.van der Aart, Q.J.M.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S59702  
A:Accession: S59716  
A:Molecule type: DNA  
A:Residues: 1-1332 <VAM>  
A:Cross-references: UNIPARC:UPI000005310C; EMBL:X76294; NID:g974203; PIDN:CAAS3940.1; PID:A.Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:SPT7  
A:Cross-references: SGD:S0000285; MIPS:YBR081C  
A:Map position: 2R  
C:Keywords: nucleus; transcription regulation  
F:466-521/Domain: bromodomain homology <BRO>

Query Match 10.7%; Score 90.5; DB 2; Length 1332;  
Best Local Similarity 24.5%; Pred. No. 80;  
Matches 39; Conservative 26; Mismatches 57; Indels 37; Gaps 8;

QY 25 VTIQNGKMSSTI-VSEEDFILPVYKGELEKGYQPDGWEISGFEGKGDAGYVINI-SKDTFI 84  
Db 549 ITIRNRADLEKEI---EDM-----EKDKDYELDEEEVAGSGRKG---LNMGAHMLA 594  
QY 85 K---PVPFKLIEKKEENKPTFD-----VSKKKD-----NPQVNHSQLNESH 123  
Db 595 KENGKVSSEKSSKTVKDEAPTNDKLTSVIPEGEKSKDKTASSTVTVHENYVKNQEIENG 654  
QY 124 RKEDLQR--EHSQKSDSTKQVATVLD--KNNISSKSTTN 160  
Db 655 KNEQDMVESSKTSKDEADAAKKTEDGLQDKTAEN 693

RESULT 40  
T28676  
rhostry protein - Plasmodium yoelii (fragment)  
C:Species: plasmodium yoelii  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
A:Accession: T28676; A45521  
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A:Title: Comparison of two members of a multigene family coding for high-molecular mass proteins.  
A:Reference number: Z20507; MUID:97077455; PMID:8920022  
A:Accession: T28676  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2401 <SIN>  
A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:g1041784; PID:R.Keen, J.; Holder, A.A.; Playfair, J.; Lockyer, M.; Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A:Title: Identification of the gene for a Plasmodium yoelii rhostry protein. Multiple clones.  
A:Reference number: A45521; MUID:91101660; PMID:2270106  
A:Accession: A45521  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2260-2401 <KEE>  
A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

Query Match 10.7%; Score 90.5; DB 2; Length 2401;

Best Local Similarity 27.4%; Pred. No. 1.5e+02;  
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;

QY	26	TIQNGKMSSTI-----VSEDFILPVYKGELEKGY--OPDGWEISGFEGKXDA	72
Db	924	TSKNHEEKISDIRKNSLKIIQDFSSESYINDI-KKELEKNVLESQNNNTDINQYLSKIEN	982
QY	73	GYVINLSKDTFIKPVFKKIEKKBEENKPTPDVSKKONPQVNHSQLNESHKEDLQREE	132
Db	983	IY--NILKLNKIKKIIDKVKYTDIEK-----NNKKINAELSNSEKIITQLKENSLLKE	1035
QY	133	HSQKSDSTKD-----VTATVLDKNNISS	155
Db	1036	COSKIKSTIDDNVSECINKNITNLKTYIVNEKNNT	1072

Search completed: April 24, 2006, 15:01:16  
Job time : 18.2438 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 12.8412 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773  
Perfect score: 848  
Sequence: 1 TTVKEFILKDTGEVSELKP.....ATVLDKNNISSKSTNNPNK 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.New.\*  
1: /SIDSS/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /SIDSS/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /SIDSS/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /SIDSS/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
5: /SIDSS/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
6: /SIDSS/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /SIDSS/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /SIDSS/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	14.0	746	US-10-793-626-652	Sequence 652, App
2	111.5	13.1	354	US-11-189-817-2	Sequence 2, Appli
3	110.5	13.0	651	US-11-128-660-1	Sequence 1, Appli
4	95	11.2	886	US-11-087-099-11456	Sequence 11456, A
5	93.5	11.0	501	US-10-485-517-381	Sequence 381, App
6	93.5	11.0	645	US-10-485-517-244	Sequence 244, App
7	92	10.8	299	US-10-793-626-1888	Sequence 1888, App
8	87	10.3	140	US-11-096-568A-4771	Sequence 4771, Ap
9	86	10.1	443	US-11-188-298-1015	Sequence 1015, Ap
10	86	10.1	700	US-11-196-475-74	Sequence 74, Appl
11	85.5	10.1	734	US-11-072-512-2272	Sequence 2272, Ap
12	85	10.0	472	US-10-793-626-658	Sequence 658, App
13	85	10.0	493	US-11-096-568A-3070	Sequence 3070, Ap
14	85	10.0	493	US-11-096-568A-3071	Sequence 3071, Ap
15	85	10.0	510	US-11-096-568A-3069	Sequence 3069, Ap
16	85	10.0	636	US-10-485-517-170	Sequence 170, App
17	85	10.0	700	US-11-196-475-66	Sequence 66, Appl
18	85	10.0	708	US-11-196-475-76	Sequence 76, Appl
19	85	10.0	720	US-10-793-626-2058	Sequence 2058, Ap
20	85	10.0	785	US-10-793-626-264	Sequence 264, App
21	84.5	10.0	1694	US-11-052-554A-83	Sequence 83, Appl
22	83.5	9.8	248	US-10-793-626-464	Sequence 464, App
23	83.5	9.8	635	US-10-821-234-1573	Sequence 1573, Ap
24	83.5	9.8	732	US-10-469-469-318	Sequence 318, App
25	83.5	9.8	732	US-11-187-230-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1

US-10-793-626-652  
; Sequence 652, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 652  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-652

Query Match	14.0%	Score 119;	DB 6;	Length 746;
Best Local Similarity	27.0%	Pred. No. 0.031;		
Matches	47;	Conservative 24;	Mismatches 57;	Indels 46; Gaps 9;
Qy	10	KDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEDFILPVYK-----GELEKGYQFDGW--	61	
Db	585	EDSVNAOSLKP-----ITIGNKQIKQOSVSGTKVLPHSKVLMTDGSLTWP-DMTGWTK	639	
Qy	62	-EISGFE-----GKKDAGYVIN--LSKDTFIKVPFKIKBEKEENKPTFDVS----K	107	
Db	640	EDVLAFEDLTKLKYSTKNGFVNTQSIKQIILK-----NDKIEVLSLRAED	686	
Qy	108	KDKNPQVNHSQLNSHRKEDLQREHSQKSDSTKDVATVLDKNNISKSSTNN	161	
Db	687	TDDQEKTDDESSDNKSKDKADEHNTSSTKN-----DKSNADSKNDSDD	734	

RESULT 2

US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US20060030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSITUUT PASTEUR  
; APPLICANT: DRULIHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM

```
; TITLE OF INVENTION: MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES
; FILE REFERENCE: 275601US0
; CURRENT APPLICATION NUMBER: US/11/189,817
; CURRENT FILING DATE: 2005-07-27
; PRIOR APPLICATION NUMBER: 60/598,062
; PRIOR FILING DATE: 2004-08-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-189-817-2

Query Match          13.1%; Score 111.5; DB 7; Length 354;
Best Local Similarity 22.5%; Pred. No. 0.052; Mismatches 62; Indels 45; Gaps 7;
Matches 41; Conservative 34;

Qy 1 TTVKEFILM-KDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFD 59
Db 154 TTKTEVAEKAKAYEKKANQYKQAVLKAEASS-----YDYIL----- 194

Qy 60 GWEISGF--EGKKDAG-----YVINLSKDTFFKVPFKIEKKER-----EN 99
Db 195 GWEFGGVPHEKKENMLSHLYSSKKKENISKENDVDLKEKEEAEETEELBEKNEE 254

Qy 100 KPTFDVSKKDKPNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTT 159
Db 255 ETEISEISEDEEEEBEKEERKKQKQKQSNENNDDQKDMAE-----QNLISKNN 309

Qy 160 NN 161
Db 310 NN 311

RESULT 3
US-11-128-660-1
; Sequence 1, Application US/11/128660
; Publication No. US20060024324A1
; GENERAL INFORMATION:
; APPLICANT: Statens Serum Institut
; TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plasmo
; FILE REFERENCE: 15007dk
; CURRENT APPLICATION NUMBER: US/11/128,660
; CURRENT FILING DATE: 2005-05-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-128-660-1

Query Match          13.0%; Score 110.5; DB 7; Length 651;
Best Local Similarity 22.8%; Pred. No. 0.13; Mismatches 38; Indels 59; Gaps 10;
Matches 46; Conservative 38;

Qy 4 KEFILM-KDTGEVSELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 53
Db 422 EEAVSEKNAHETVE---HEETVQESNPEKADNDGNVQSNNELNENEV-----ESE 472

Qy 54 K-----GQFPD-GWEISGF--EGKKDAG-----YVINLSKDTFFKVPFKIE 92
Db 473 KSEHEARSKAKEASSYDYLGMWFGGVPHEKKENMLSHLYSSKKKENISKENDVDVLD 532

Qy 93 EKKEENKPTFDVSKKDKPNQVNHSQLN-----ESHKEDLQREHSQKSDS 139
Db 533 E-KEEAEETEELBEKNEEETEISEISEDEEEEBEKEERKKQKQKQSNENNDDQ 591

Qy 140 TKDVTATVLDKNNISKSTTN 161
Db 592 KDMAE-----QNLISKNNNN 608
```

## RESULT 4

```
US-11-087-099-11456
; Sequence 11456, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 11456
; LENGTH: 886
; TYPE: PRT
; ORGANISM: Bacillus cereus ATCC 14579
US-11-087-099-11456

Query Match          11.2%; Score 95; DB 7; Length 886;
Best Local Similarity 24.5%; Pred. No. 3.8; Mismatches 75; Indels 48; Gaps 10;
Matches 49; Conservative 28;

Qy 8 LNKDTGEVSELKP-----HRVTVTIQNGKEMSTIVSEEDFILPVY--KGE----- 51
Db 281 INLGANEIFDVRPVEELGKRISIDIQRKIFLNEASVDEEIKIPVYNLKGEPLOQINLQS 340

Qy 52 ----LEKGYQFDGWEISG---FEGKKDAGYVINLSKDTFFKVPFKIEKKEENKPTFD 104
Db 341 EGATLNNG--FIRWNSPGEKIYEPKLDTN--STESIRFNGTVIQTIVIVEKQESQNVLD 396

Qy 105 VSKKK--DNQVNHSQLNESHKEDL-----QREHSQKSDSTKDVTA---TVLDK---- 150
Db 397 KTLQOHINKENLGRNVNVPITKEDLLQIKLEILKEKNEIKDITGLEVTNLEKLTLE 456

Qy 151 -----NNISSKSTTNPN 163
Db 457 GVGLKNIEFISNLKQLNNVN 476
```

## RESULT 5

```
US-10-485-517-381
; Sequence 381, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-381

Query Match          11.0%; Score 93.5; DB 6; Length 501;
Best Local Similarity 21.9%; Pred. No. 2.5;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

Qy 1 TTVKEFILM-KDTGEVSELKPHRVTV--TIQNGKEMSTIVSEEDFILPVYKGELE-KGY 56
Db 183 SAITEFQNVQPTNEKMTDLDQTKVYVESVENNESMDTFVKH-----PIKTGMLNGKKY 237
```

QY 57 QF-----DGEISGPEGK-----D 71  
Db 238 MVMETNDYKDFWVEGQRTISKAKNTRTIIPYVEGKTYDAIVKVHVKTIDYD 297  
QY 72 AGVINL-SKDTFIKPVFKKIEKKEENKPTFDV-----SKKDNPOVNHSQLNESHK 125  
Db 298 GQYHVIQDKAFTKANTDKSKGQDNSAKKEATPATPSKPTSPVEKESQKDSQKD 357  
QY 126 EDLQ-----REHSQKSDSTQV--ATVLDKNNISSKSTNNPNK 164  
Db 358 DNKQLPSVEKENDASSESGDKTPTKPKGEVSSSTI--PTK 399  
RESULT 6  
US-10-485-517-244  
; Sequence 244, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, Simon  
; APPLICANT: Mond, James  
; TITLE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629W0  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; CURRENT FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118825.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 244  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-244

Query Match 11.0%; Score 93.5; DB 6; Length 645;  
Best Local Similarity 21.9%; Pred. No. 3.4;  
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;  
QY 1 TVVKEFILNDTGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 56  
Db 327 SAITEFQNVQPTNERQMTLDQTKYVYVESVENNESMDTFVKH-----PIKTGLMGKKY 381  
QY 57 QF-----DGEISGPEGK-----D 71  
Db 382 MVMETNDYKDFWVEGQRTISKAKNTRTIIPYVEGKTYDAIVKVHVKTIDYD 441  
QY 72 AGVINL-SKDTFIKPVFKKIEKKEENKPTFDV-----SKKDNPOVNHSQLNESHK 125  
Db 442 GQYHVIQDKAFTKANTDKSKGQDNSAKKEATPATPSKPTSPVEKESQKDSQKD 501  
QY 126 EDLQ-----REHSQKSDSTQV--ATVLDKNNISSKSTNNPNK 164  
Db 502 DNKQLPSVEKENDASSESGDKTPTKPKGEVSSSTI--PTK 543

RESULT 7  
US-10-793-626-1888  
; Sequence 1888, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P03480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 1888  
; LENGTH: 299  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-1888

Query Match 10.8%; Score 92; DB 6; Length 299;  
Best Local Similarity 24.4%; Pred. No. 1.8;  
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;  
QY 9 NKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFI-----LPVYKGELEK--YQPDGWEI 63  
Db 33 NKDT-EKSDKKYHRIISLIPNTBILYRLGIGEDIVGVSTVDDYPKDKVKGKQKQFDAMNL 91  
QY 64 SGPE-----GKKDAGVYVNLSDTKPKV-----FKKIEKK 95  
Db 92 NKBELIKAKPDLILAHESQKNSAGKVLKSLDKGVYVYVVKDAQSIDETDTFKSIGQLT 151  
QY 96 EENKPTFDVSKKDNPOVNHSQLNESHKEDLQREHSQ 135  
Db 152 DREKQAKELVDTEKHVKEKIINSVPKHKKQEVPMVEVSSK 191

RESULT 8  
US-11-096-568A-4771  
; Sequence 4771, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 4771  
; LENGTH: 140  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(140)  
; OTHER INFORMATION: Ceres Seq. ID no. 14304111  
US-11-096-568A-4771

Query Match 10.3%; Score 87; DB 7; Length 140;  
Best Local Similarity 25.9%; Pred. No. 1.8;  
Matches 36; Conservative 28; Mismatches 43; Indels 32; Gaps 7;  
QY 30 GKMSSTI-----VSEEDFILPVY--YGELEKGYQPDGWEISGPEGKDGAGYVNLSDT 82  
Db 8 GKQGSASAREDHGSGEEDKIPAYRRGRPKQPKMD-----FEEDEDEB----- 52  
QY 83 FIKPVFKKIEKKEENKPTFDVSKKDNPO-----VNHSQLNESHKEDLQREHSQKSD 138  
Db 53 --BELVKEEEEEEDDSVT---SKKENERKRWINGSNTDANEKENGLOKSKSRDSS 107  
QY 139 STKDVATVLDKNNISSK 157  
Db 108 STKS-TSTGFRQNGSRKS 125

RESULT 9  
US-11-188-298-1015  
; Sequence 1015, Application US/11188298  
; Publication No. US2006007522A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53452)B

```
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 1015
; TYPE: PRT
; ORGANISM: Thermococcus kodakaraensis
US-11-188-298-1015

Query Match          10.1%; Score 86; DB 7; Length 443;
Best Local Similarity 17.8%; Pred. No. 9;
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

QY 48 YKELEKGYQPCGWEISGEGKKDAGYVNLKDTFPI-----84
Db 42 YEAEVEDGVSFGSGIPGEGIEDSLIFKADPSTYAEIPWEGIGRVGYIYKGDPEYQA 101
QY 85 --KPVFKIEEKEEEN-----KPTDVSKKD-----110
Db 102 DPGILKRVLEKEKEGLKAHIGPEPFYIFKNGTWELHPDSCGYFDLVGLDKARIR 161
QY 111 -----NPQVNSQLNESHKEDLQREHSQKSD-----STKDVATATVLD 149
Db 162 REIALYMPYLGLKPEVLHHEVSKAQHEIDFRYDEALRTADNIVSPKVVYKVAE 215

RESULT 10
US-11-196-475-74
; Sequence 74, Application US/11/196,475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: Fastseq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-74

Query Match          10.1%; Score 86; DB 7; Length 700;
Best Local Similarity 23.2%; Pred. No. 16;
Matches 41; Conservative 38; Mismatches 54; Indels 44; Gaps 9;

QY 23 VTTTQNGKMSSTIVSEEDFILPVYKGELEKGYQD-----GWEISGEGKK 70
Db 258 ITETIENLADQEKATDEE-----HKKEIES--QVDAKKQKEELDKAIDLAKAOKKL 309
QY 71 D-AGYVNLKDTFTFKPVFKIEEKEEENKEP-TFDVSKKNPNV-----HSOLN 120
Db 310 DPAENLDIQRTVREKQENINETNKEKNLPKPGDVSPKVDKQLQKESLEDLQEQK 369
QY 121 ES---HRKEDLQREHSQKSD-----STKDVATVLDKNNISSKSTTNNPNK 164
```

```
Db 370 ETGDENQRIETEKQIEIKKSKDEKLLKSKODKASKDGKALDLDR-ELNSKASSKEKSK 425

RESULT 11
US-11-072-512-2272
; Sequence 2272, Application US/11/072,512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2272
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2272

Query Match          10.1%; Score 85.5; DB 7; Length 734;
Best Local Similarity 21.7%; Pred. No. 18;
Matches 39; Conservative 42; Mismatches 58; Indels 41; Gaps 11;

QY 10 KDTG-EVSELKPHRVTV-----TIQNGKMSSTIVSEE--DFILPVYK 49
Db 168 EDTGIYVSEVDPNSIAAKDGRIGEDVQNGREAVALLSNDECKRIVLLVAR 227
QY 50 GELEKGYQPCGWEISGEGKKDAGYVNLKDTF-----IKPVFKIEE-KKEEENKP 101
Db 228 PEIQLD---EGW-----LEDERNE-FLEELNLEMLEEHEEHEAMQPTANEVEQPKQEEEG 279
QY 102 TFDVSKKNPNVNSQLNESHK-KEDLQREHSQKSDSTKDVATATVL-DKNNISSKSTT 159
Db 280 TTDATSSN---NHEKDSGVGRTDESURDESSEQENAAEDPNSTLSKSRDLGQSQDT 336

RESULT 12
US-10-793-626-658
; Sequence 658, Application US/10/793,626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
```





```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-66

Query Match      10.0%; Score 85; DB 7; Length 700;
Best Local Similarity 24.4%; Pred. No. 19;
Matches 44; Conservative 33; Mismatches 53; Indels 50; Gaps 9;

QY   23 VVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYPDQGWISGFEFGKDAGYYVNLK-- 80
DB   258 ITETIENLRDQLEKATDEE-----HKKEIES--QVDAKKKQKEELDCKA---INLDKAQ 306
QY   81 -----DTFTKPVPFKIEEKKEBENKP-TFDVSKKDNPKVN-----HS 117
DB   307 QKLDSAEADNLDVQRNTVREKIQEDINEINKSNLFPKPDVSSPKVDQLQIKESLEDLOE 366
QY   118 QLNES---HRKEDIQREHRSQSD-----STKDVTTATVLDDKNNISKSSTNNPNK 164
DB   367 QLKETGDENQKRETEKQIEIKKSDKLLKSKDKDKASKDGKALDLDR-ELNSKASSKEKSK 425

RESULT 18
US-11-196-475-76
; Sequence 76, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-76

Query Match      10.0%; Score 85; DB 7; Length 708;
Best Local Similarity 27.6%; Pred. No. 19;
Matches 43; Conservative 22; Mismatches 53; Indels 38; Gaps 8;

QY   23 VVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYPDQGWISGFEFGKDAGYYVNLK-- 80
DB   258 ITETIENLRDQLEKATDEE-----HKKEIES--QVDAKKKQKEELDCKA---INLDKAQ 306
QY   81 -----DTFTKPVPFKIEEKKEBENKP-TFDVSKKDNPKVN-----HS 117
DB   307 QKLDSAEADNLDVQRNTVREKIQEDINEINKSNLFPKPDVSSPKVDQLQIKESLEDLOE 366
QY   126 EDLORE----EHSQKSDSTKDVTTATVLDDKNNISK 156
DB   362 EDLQOLKEAGDENQKREIEKQIEIKRDEELLKSK 397

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-66

Query Match      10.0%; Score 85; DB 6; Length 636;
Best Local Similarity 22.0%; Pred. No. 17;
Matches 35; Conservative 35; Mismatches 61; Indels 28; Gaps 7;

QY   7 ILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYPDQGWISGF 66
DB   27 IVTKDYSGKSVNQAGSKNGTLDRLSYLSALYLEDI--IYAIGLTNKEYE-----GD 78
QY   67 EGKKGAGYVINLSKDTFFIKPVEFK-----IEEKKEE-ENKPTFDVSKKDNPKVNHSQLN- 120
DB   79 NIYKEA-----KORLKEVLQOYLKERKKSQYEDYKQWYANYKKENPRTLRKMANF 131
QY   121 ESHRKEDLQREHRSQSDSKTDV-----TATVLDDKN 152
DB   132 HKYNLEELSWEYNELODKALKALDDFHREVKDIKXNS 170

RESULT 17
US-11-196-475-66
; Sequence 66, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
```





; OTHER INFORMATION: Ceres Seq. ID no. 2712010  
US-11-096-568A-28315

Query Match 9.8%; Score 83.5; DB 7; Length 1036;  
Best Local Similarity 24.1%; Pred. No. 41;  
Matches 41; Conservative 32; Mismatches 72; Indels 25; Gaps 7;

QY 1 TTVKEFILNK--DTGEVS-----ELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGE 51  
DB 492 TSSNKKIVKVAETGDTSDPSAKANEQTPAKTIIVKKIIRKVAEKVAEIDNKM---DGD 548  
QY 52 LEKGYPQDGEISGPEGKK--DAGVYINLSKDTFIKVPFKIEKKKEENKPTPDVSKKK 109  
DB 549 SKDGDSDKKVMEV-GKKSSDGSV-----EMKPTAESLEDVKDENASKTVDVKQET 600  
QY 110 DNPQVNHSQLNESHKEDLQREH---SQKSDSTKDVATVLDKNNISSK 156  
DB 601 GSPDTKKKGAGSSSSKKDTKTGDKKAEKKNNSETMSEGGKKIDRNTDEK 650

## RESULT 27

US-11-096-568A-28314  
; Sequence 28314, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28314  
; LENGTH: 1070

; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1070)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712009  
US-11-096-568A-28314

Query Match 9.8%; Score 83.5; DB 7; Length 1070;  
Best Local Similarity 24.1%; Pred. No. 43;  
Matches 41; Conservative 32; Mismatches 72; Indels 25; Gaps 7;

QY 1 TTVKEFILNK--DTGEVS-----ELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGE 51  
DB 526 TSSNKKIVKVAETGDTSDPSAKANEQTPAKTIIVKKIIRKVAEKVAEIDNKM---DGD 582  
QY 52 LEKGYPQDGEISGPEGKK--DAGVYINLSKDTFIKVPFKIEKKKEENKPTPDVSKKK 109  
DB 583 SKDGDSDKKVMEV-GKKSSDGSV-----EMKPTAESLEDVKDENASKTVDVKQET 634  
QY 110 DNPQVNHSQLNESHKEDLQREH---SQKSDSTKDVATVLDKNNISSK 156  
DB 635 GSPDTKKKGAGSSSSKKDTKTGDKKAEKKNNSETMSEGGKKIDRNTDEK 684

## RESULT 28

US-11-096-568A-28313  
; Sequence 28313, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28313  
; LENGTH: 1276

; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(1276)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712008  
US-11-096-568A-28313

Query Match 9.8%; Score 83.5; DB 7; Length 1276;  
Best Local Similarity 24.1%; Pred. No. 54;  
Matches 41; Conservative 32; Mismatches 72; Indels 25; Gaps 7;

QY 1 TTVKEFILNK--DTGEVS-----ELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGE 51  
DB 732 TSSNKKIVKVAETGDTSDPSAKANEQTPAKTIIVKKIIRKVAEKVAEIDNKM---DGD 788  
QY 52 LEKGYPQDGEISGPEGKK--DAGVYINLSKDTFIKVPFKIEKKKEENKPTPDVSKKK 109  
DB 789 SKDGDSDKKVMEV-GKKSSDGSV-----EMKPTAESLEDVKDENASKTVDVKQET 840  
QY 110 DNPQVNHSQLNESHKEDLQREH---SQKSDSTKDVATVLDKNNISSK 156  
DB 841 GSPDTKKKGAGSSSSKKDTKTGDKKAEKKNNSETMSEGGKKIDRNTDEK 890

## RESULT 29

US-11-096-568A-17853  
; Sequence 17853, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 17853  
; LENGTH: 305

; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(305)  
; OTHER INFORMATION: Ceres Seq. ID no. 12361233  
US-11-096-568A-17853

Query Match 9.7%; Score 82.5; DB 7; Length 305;  
Best Local Similarity 19.6%; Pred. No. 11;  
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;

QY 20 PHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDGAGVYINLS 79  
DB 40 PDLVDCSLNAGDAGSSKKKAEKSSFRPVAK-----ETPSLEDSEKKKTQKAS 87  
QY 80 KDTFIKVPFKIEKKKEENKPTPDVSKKKDNPNQVNHSQLNESHKEDL 128  
DB 88 NQHSVK---KDIEESNESVVKPQVGVSTPSYGFSGFKCDERAEKREFYFYSKLEEKIHAQEL 144  
QY 129 QREHESQKSDSTKDVATVLDKN-----TFDVSKKKDNPNQVNHSQLNESHKEDL 151

DB 145 EKSNIQAKSKETBEAEKQKLRKSLNFKATPMPSPYKPEPPPKVELKKIPTTTRASPGLGR 204  
QY 152 --NISSKSTTNPN 163  
DB 205 SKNTSSGGTGNPN 218

## RESULT 30

US-11-096-568A-17852  
; Sequence 17852, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

```
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17851
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; NAME/KEY: misc_feature
; LOCATION: (1)..(327)
; OTHER INFORMATION: Ceres Seq. ID no. 12361232
US-11-096-568A-17851

Query Match
Best Local Similarity 9.7%; Score 82.5; DB 7; Length 327;
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;

QY 20 PHRVTVTIQNGKEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFGEGKKDAGVYNLS 79
DB 62 PDLVDCSLNGDAGSSKKAERSSRPVAK-----ETPSLEDSNEKKKTQKAS 109
QY 80 KDTFKVPVKKIEEKKENKP-----TFDVSKKKDNPOVNHSQLNESHKEDL 128
DB 110 NQHSVK---KDIEESNESVKPQRVGSTPSYGFSPKCDERAEKRREFYSKLEEKIHAQEL 166
QY 129 QREHSQKSDSKQVATVLDKN-----TFDVSQKKNPOVNHSQLNESHKEDL 151
DB 167 EKSNLQAKSETEAEELKRLKSLNFKATPMPSPFYKEPPPKVELKKTPTTRARSPKLGR 226
QY 152 --NISKSTNNPN 163
DB 227 SKNTSSGGTGNPN 240

RESULT 31
US-11-096-568A-17851
; Sequence 17851, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 17851
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; NAME/KEY: misc_feature
; LOCATION: (1)..(386)
; OTHER INFORMATION: Ceres Seq. ID no. 12361231
US-11-096-568A-17851

Query Match
Best Local Similarity 9.7%; Score 82.5; DB 7; Length 386;
Matches 38; Conservative 22; Mismatches 69; Indels 65; Gaps 4;

QY 20 PHRVTVTIQNGKEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFGEGKKDAGVYNLS 79
DB 121 PDLVDCSLNGDAGSSKKAERSSRPVAK-----ETPSLEDSNEKKKTQKAS 168
QY 80 KDTFKVPVKKIEEKKENKP-----TFDVSKKKDNPOVNHSQLNESHKEDL 128
DB 169 NQHSVK---KDIEESNESVKPQRVGSTPSYGFSPKCDERAEKRREFYSKLEEKIHAQEL 225
QY 129 QREHSQKSDSKQVATVLDKN-----TFDVSQKKNPOVNHSQLNESHKEDL 151
```

```
Db 226 EKSNLQAKSETEAEELKRLKSLNFKATPMPSPFYKEPPPKVELKKTPTTRARSPKLGR 285
QY 152 --NISKSTNNPN 163
DB 286 SKNTSSGGTGNPN 299

RESULT 32
US-11-087-099-5657
; Sequence 5657, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 5657
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Bacillus anthracis str. Ames
US-11-087-099-5657

Query Match
Best Local Similarity 9.7%; Score 82.5; DB 7; Length 1070;
Matches 37; Conservative 24; Mismatches 66; Indels 27; Gaps 6;

QY 12 TGEVSELKP-----HRVTVTIQNGKEMSTIVSEDFILPVY-----KGELEKGY 56
DB 411 TNEIRDIPQELGKRIKIDAQKQVFLDEAYMNEEVKIPYVDYNGTALQNIWKSEGG 470
QY 57 QFDG---WEISG---FEGKKDAGVYNLSKDTFKIPVKFKIEEKKENKPTFDVSKKK- 109
DB 471 ITNGVIKWSLGEKWEYFKMDAGE---SKIRFGQRVIONIVEKRESSNVIQDKMLRQY 526
QY 110 -DNQVNHSQLNESHKEDLQREHSQKSDSKD 142
DB 527 MNKHNFERNVNTPTITKEDLLTVKALKITDGKCE 560

RESULT 33
US-11-096-568A-4567
; Sequence 4567, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4567
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc_feature
; LOCATION: (1)..(339)
; OTHER INFORMATION: Ceres Seq. ID no. 15219423
US-11-096-568A-4567

Query Match
Best Local Similarity 9.7%; Score 82; DB 7; Length 339;
Matches 43; Conservative 32; Mismatches 56; Indels 36; Gaps 9;

QY 9 NKDTGSEVSELKPHRVTVTIQNGKEMSTIVSE-----EDFILPVYK---GELEKGY 55
DB 114 NGKTDGVE--RPRRV-YDRRSSTGRSDNVKREGGGRGNWGTPTDQIQVTEPTTEVEKS 170
QY 56 --YQFDGWEISGFGEGKKDAGVIN--LSKDTFKIPVKFKIEEKK-----EENK---P 101
```

Db 171 PVAKEGGEDATTDAKGEAPEVEQEPEDKEMTLEBYEKLEBKKAQAQTKVEERKVDTK 230  
QY 102 TFDVSKKDNQPNVHSQL-----NESHKEDLQREHSQKSDSTKD 142  
Db 231 VPESMQLSNKKTNDSEIFIKLGSDDKDKDAAEKEEKAQKSLSINE 277

## RESULT 34

US-11-188-298-15964  
; Sequence 15964, Application US/11188298  
; Publication No. US20060075522A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53452)B  
; CURRENT APPLICATION NUMBER: US/11/188,298  
; CURRENT FILING DATE: 2005-07-22  
; PRIOR APPLICATION NUMBER: 60/592,978  
; PRIOR FILING DATE: 2004-07-31  
; NUMBER OF SEQ ID NOS: 22569  
; SEQ ID NO 15964  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: *Pyrococcus furiosus* DSM 3638  
US-11-188-298-15964

Query Match 9.7%; Score 82; DB 7; Length 439;  
Best Local Similarity 18.1%; Pred. No. 19;  
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;  
QY 55 GYQPDGWSISGPEGKDGAGYVNLKDTFI-----KP-----VFK 89  
Db 45 GISPDGSSVPGFGIEDSLVFKADPDITVEVPMDNVARVGYFYKKNKPYGADPRGILK 104  
QY 90 KIEEKKKEEN-----KPTEDVSKKKN-----111  
Db 105 RALSELEKEGYKAYIGPPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIABYM 164  
QY 112 -----PQVNHSQLNESHKEDLQREHSQKSD---STKDVATATVLDKNNI 153  
Db 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKATADNIVSPKYITKAVAEHGL 215

## RESULT 35

US-11-188-298-16606  
; Sequence 16606, Application US/11188298  
; Publication No. US20060075522A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT  
; FILE REFERENCE: 38-21(53452)B  
; CURRENT APPLICATION NUMBER: US/11/188,298  
; CURRENT FILING DATE: 2005-07-22  
; PRIOR APPLICATION NUMBER: 60/592,978  
; PRIOR FILING DATE: 2004-07-31  
; NUMBER OF SEQ ID NOS: 22569  
; SEQ ID NO 16606  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: *Pyrococcus woesei*  
US-11-188-298-16606

Query Match 9.7%; Score 82; DB 7; Length 439;  
Best Local Similarity 18.1%; Pred. No. 19;  
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;  
QY 55 GYQPDGWSISGPEGKDGAGYVNLKDTFI-----KP-----VFK 89  
Db 45 GISPDGSSVPGFGIEDSLVFKADPDITVEVPMDNVARVGYFYKKNKPYGADPRGILK 104  
QY 90 KIEEKKKEEN-----KPTEDVSKKKN-----111

Db 105 RALSELEKEGYKAYIGPPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIABYM 164  
QY 112 -----PQVNHSQLNESHKEDLQREHSQKSD---STKDVATATVLDKNNI 153  
Db 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKATADNIVSPKYITKAVAEHGL 215

## RESULT 36

US-10-793-626-2482  
; Sequence 2482, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P03480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2482  
; LENGTH: 568  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-2482

Query Match 9.7%; Score 82; DB 6; Length 568;  
Best Local Similarity 23.2%; Pred. No. 26;  
Matches 38; Conservative 28; Mismatches 54; Indels 44; Gaps 7;  
QY 8 LNKDTGEVSELKPHRVTVTIQ-----NGKEMSSTIVSEEDFILPVYKGELEKGYQ 57  
Db 199 LNEATTLOEKVNOQVQPLREDMSFLTRFNKHAHVEKILNQEDYDVVPSLISEL---YQ 255  
QY 58 PDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIEEKKKEENKPTFDVSKKKNQPNVHNS 117  
Db 256 TQ-TSISN-----TIVL-----TFSDEVIKALKKLENESTPLFEEA-----VNHV 295  
QY 118 QLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISKSTNN 161  
Db 296 QVNEISSLSDENEDRYEYDR-----YIELNTLKDSLTSN 328

## RESULT 37

US-10-475-204-34  
; Sequence 34, Application US/10475204  
; Publication No. US20050277116A1  
; GENERAL INFORMATION:  
; APPLICANT: PRESIDENT AND FELLOWS OF HARVARD COLLEGE  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE IDENTIFICATION OF  
; TITLE OF INVENTION: PROTEIN INTERACTIONS IN VERTEBRATE CELLS  
; FILE REFERENCE: HMV-056.25  
; CURRENT APPLICATION NUMBER: US/10/475,204  
; CURRENT FILING DATE: 2003-10-17  
; PRIOR APPLICATION NUMBER: PCT/US02/13008  
; PRIOR FILING DATE: 2002-09-27  
; PRIOR APPLICATION NUMBER: 60/285,509  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 34  
; LENGTH: 943  
; TYPE: PRT  
; ORGANISM: *Homo sapiens*  
US-10-475-204-34

Query Match 9.7%; Score 82; DB 6; Length 943;  
Best Local Similarity 23.2%; Pred. No. 49;  
Matches 44; Conservative 18; Mismatches 76; Indels 52; Gaps 7;

Qy	13	GEVSELKPHRV--TVTIQNGKMSSTVISEEDFIILPVYKGELEKGYQPDGWEISGPGCKG	71
Db	343	GRKSREKHNILPKTLANDKHSKHPVETS--QPSDKTVLDTSYALIDETVNNYRSTKY	400
Qy	72	AGVIVNLSDKDTIKPVFKKIEKK-----EENKPTFDVSKKD-NPQVNHSQLNESHK	124
Db	401	EMYSKNAEKPSRSKRTIKQKQRKFMKPAEQ----LDVGQSKDENIHTSHITQDEFOR	456
Qy	125	KEDLOREEH-----SOKSDSTKD-----VTAT	146
Db	457	NSDRNMEEHGNDVCVKQMPVGSKKSTRKDKESKKGKGFSSBKNLVPEEVTST	516
Qy	147	VLDKNNISK	156
Db	517	VTKSRRISSR	526

```

RESULT 38
US-10-793-626-1432
; Sequence 1432, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1432
; LENGTH: 1145
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1432

```

Query Match	9.7%	Score 82;	DB 6;	Length 1145;
Best Local Similarity	23.2%;	Pred. No. 62;		
Matches	38;	Conservative 28;	Mismatches 54;	Indels 44; Gaps 7;
Qy	8	LAKDTGEVSELKPHRVTVTIQ-----NGKEMSTIVSEBDFILPVYKGLEKGY	57	
Dd	377	LNEATTLQLEKNQVOQLREDMSFLTRFKINHAVNEKI LNQEYDVVPISLSEL---	433	
Qy	58	PDGWEISGFEGKCDAGYYINLSKDTFIKPVPFKCIEEKEEENKTFDVSKKONPQNHS	117	
Dd	434	TQ-TSISN-----TYVL-----TFSDEVIKALNKKIENESTPLFEEA-----VNHV	473	
Qy	118	QLNSHREKEDLQRHEHSOKSDSKDVTATVLDKNNISSKTNN	161	
Dd	474	QWNELSSDENEDRYEYDR-----YTALNTLKDSLTSHN	506	

```

RESULT 39
US-11-096-568A-32043
; Sequence 32043, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 32043
; LENGTH: 313
; TYPE: PRT

```

```

; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(313)
;
; OTHER INFORMATION: Ceres Seq. ID no. 13592023
US-11-096-568A-32043

Query Match          9.6%; Score 81.5; DB 7; Length 313;
Best Local Similarity 21.6%; Pred. No. 14;
Matches 32; Conservative 28; Mismatches 63; Indels 25; Gaps 5;

Qy 17 ELKPHRVTVTIQNGKEMSSITVSECC-----FILPVYK-----GELEKGYOFDGGWEIS 64
    ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| ::
Db 166 EWPKTKIKNPAYGKWKAPMIDNPFEKDDDELIVFPKLVGVELWQVKGSLFDNVLVIS 225
    ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| ::
Qy 65 GFEGCKDAGVIVILSKDTF-----IKPVFKKIEKKKEENKPTFDVYSKKKNPQNVNHSQ 118
    ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| ::
Db 226 -----DPEYAKLAETWGHKDKAACKAADEAKKKEEESK--DAPAESDAEEEAEDD 278
    ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| :: ||| ::
Qy 119 LNESHRKEDLQREHSHQKSDSTKDVTAT 146
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 279 DNEGDDSDNSKSGEETKEAETKEAET 306
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

```

RESULT 40
US-11-096-568A-32042
; Sequence 32042, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polynucleotide
; TITLE OF INVENTION: Tharby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 32042
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(425)
; OTHER INFORMATION: Ceres Seq. ID no. 13592022
US-11-096-568A-32042

```

Query Match	9.6%;	Score 81.5;	DB 7;	Length 425;
Best Local Similarity	21.6%;	Pred. No. 20;		
Matches	32;	Conservative	28;	Mismatches 63;
				Indels 25;
				Gaps 5;
Qy	17	ELKPHRVTVTIQNGKEMSSITVISED-----FILPVVK-----GELEKGYQFDGWEIS	64	
Db	278	EWKPKKIKNPAYKGKWKAPMIDNPEFFKDDPELYVFPKLKYGVGLWQVKSGLFDNVLVS	337	
Qy	65	GFECKDAGYVINLSKDTF-----IKVPFKKIBEKKEEENKPTFDVSKKKNQPNVNSHQ	118	
Db	338	-----DPEYAKLAETGTWGHKDKAKAFTDEAKKKEEESK--DAPAESDAEEAEEDD	390	
Qy	119	LNESHKREDLQREHFSQKSDSTKDVTAT	146	
Db	391	DNEGDDSNESKSEETKBAETKBAET	418	

Search completed: April 24, 2006, 15:44:51  
Job time : 14.8412 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 82.1834 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848

Sequence: 1 TTVKEFILNKDTGEVSELKP.....ATVLDKNNISSKSTNNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	773	4	US-10-067-385-8
2	848	100.0	2119	3	US-09-769-744A-28
3	848	100.0	2140	4	US-10-282-122A-73670
4	848	100.0	2140	5	US-10-472-928-1180
5	845	99.6	637	5	US-10-617-320-3169
6	615	72.5	117	3	US-09-765-272-68
7	615	72.5	117	6	US-11-106-649-68
8	119	14.0	778	4	US-10-724-972A-5663
9	111.5	13.1	775	4	US-10-282-122A-70721
10	111	13.1	707	4	US-10-282-122A-52942
11	110.5	13.0	647	5	US-10-691-672A-3
12	108	12.7	188	5	US-10-691-672A-7
13	107.5	12.7	470	5	US-10-739-930-6262
14	106	12.5	665	3	US-09-820-843A-107
15	104	12.3	540	5	US-10-732-923-22820
16	103.5	12.2	169	5	US-10-691-672A-2
17	103	12.1	903	4	US-10-282-122A-52328
18	101.5	12.0	564	6	US-11-097-143-12723
19	101.5	12.0	948	5	US-10-732-923-4286
20	101	11.9	973	5	US-10-732-923-18783
21	100.5	11.9	1373	5	US-10-732-923-16976
22	100	11.8	3127	5	US-10-732-923-22588
23	99.5	11.7	1384	4	US-10-473-576-22
24	99.5	11.7	1404	4	US-10-473-576-2
25	97.5	11.5	1184	4	US-10-282-122A-53254
26	97.5	11.5	2468	4	US-10-755-889-615
27	97.5	11.5	2468	5	US-10-489-740-216

28	97.5	11.5	2519	5	US-10-450-763-46995	Sequence 46995, A
29	96.5	11.4	898	4	US-10-425-115-205148	Sequence 205148, A
30	96.5	11.4	2060	4	US-10-381-596A-2	Sequence 2, Appli
31	96	11.3	1419	5	US-10-732-923-4285	Sequence 4285, Ap
32	95.5	11.3	1350	5	US-10-732-923-8760	Sequence 8760, Ap
33	95.5	11.3	1434	5	US-10-732-923-4235	Sequence 4235, Ap
34	94.5	11.1	3124	5	US-10-732-923-22709	Sequence 22709, A
35	93.5	11.0	645	4	US-10-282-122A-70294	Sequence 70294, A
36	93.5	11.0	645	5	US-10-470-048B-414	Sequence 414, App
37	93.5	11.0	654	4	US-10-172-502-10	Sequence 10, Appl
38	93.5	11.0	654	6	US-11-020-509-10	Sequence 10, Appl
39	93	11.0	932	4	US-10-282-122A-52510	Sequence 52510, A
40	93	11.0	869	4	US-10-437-963-12282	Sequence 122282, A
41	93	11.0	933	4	US-10-437-963-166606	Sequence 166606, A
42	93	11.0	1529	5	US-10-732-923-8762	Sequence 8762, Ap
43	92.5	10.9	645	5	US-10-470-048B-142	Sequence 142, App
44	92.5	10.9	648	4	US-10-425-115-252919	Sequence 252919, A
45	92	10.8	295	4	US-10-282-122A-70503	Sequence 70503, A

## ALIGNMENTS

### RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 848;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 7.3e-64;	Mismatches 0;	Indels 0; Gaps 0;
Matches 164;	Conservative	0;		
QY	1	TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG	60	
Db	610	TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG	669	
QY	61	WEISGFEKGKAGYVINLSKDTFKVPVKIEKKKEENKPTFDVSKKNDPNQVNHSQLN	120	
Db	670	WEISGFEKGKAGYVINLSKDTFKVPVKIEKKKEENKPTFDVSKKNDPNQVNHSQLN	729	
QY	121	ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK	164	
Db	730	ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK	773	

### RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins

```
/ FILE REFERENCE: PWC/P21122WO
/ CURRENT APPLICATION NUMBER: US/09/769,744A
/ REMAINING PRIOR APPLICATION DATA REMOVED - See File Wrapper or PALM.
/ PRIOR FILING DATE: 2001-01-26
/ PRIOR APPLICATION NUMBER: PCT/GB99/02452
/ PRIOR FILING DATE: 1999-07-27
/ PRIOR APPLICATION NUMBER: GB 9816336.3
/ PRIOR FILING DATE: 1998-07-27
/ PRIOR APPLICATION NUMBER: US 60/125329
/ PRIOR FILING DATE: 1999-03-19
/ NUMBER OF SEQ ID NOS: 196
/ SOFTWARE: Patent in Ver. 2.1
/ SEQ ID NO 28
/ LENGTH: 2119
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 848; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 2.5e-63;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVEFILNKDTGEVSELKPHRVTTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
|
|
|
Db 1922 TTVEFILNKDTGEVSELKPHRVTTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 1981
|
|
|
QY 61 WEISGPEGKDGAGVYINLSKDTFTKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLN 120
|
|
|
Db 1982 WEISGPEGKDGAGVYINLSKDTFTKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLN 2041
|
|
|
QY 121 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 164
|
|
|
Db 2042 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 2085
|
|
|

RESULT 3
US-10-282-122A-73670
/ Sequence 73670, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Lianguo
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forayth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
```

```
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ REMAINING PRIOR APPLICATION DATA REMOVED - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: Patent in version 3.1
/ SEQ ID NO 73670
/ LENGTH: 2140
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match      100.0%; Score 848; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.6e-63;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVEFILNKDTGEVSELKPHRVTTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
|
|
|
Db 1943 TTVEFILNKDTGEVSELKPHRVTTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2002
|
|
|
QY 61 WEISGPEGKDGAGVYINLSKDTFTKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLN 120
|
|
|
Db 2003 WEISGPEGKDGAGVYINLSKDTFTKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLN 2062
|
|
|
QY 121 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 164
|
|
|
Db 2063 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 2106
|
|
|

RESULT 4
US-10-472-928-1180
/ Sequence 1180, Application US/10472923
/ Publication No. US20050020813A1
/ GENERAL INFORMATION:
/ APPLICANT: CHIRON Spa
/ APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
/ TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
/ FILE REFERENCE: P026926WO
/ CURRENT APPLICATION NUMBER: US/10/472,928
/ CURRENT FILING DATE: 2003-09-26
/ PRIOR APPLICATION NUMBER: GB-0107658.7
/ PRIOR FILING DATE: 2001-03-27
/ NUMBER OF SEQ ID NOS: 4979
/ SOFTWARE: SeqWin99, version 1.03
/ SEQ ID NO 1180
/ LENGTH: 2140
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
/ FEATURE:
/ OTHER INFORMATION: serine protease, subtilase family
/ OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
/ OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
US-10-472-928-1180

Query Match      100.0%; Score 848; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.6e-63;
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVEFILNKDTGEVSELKPHRVTTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60
|
|
|
Db 1943 TTVEFILNKDTGEVSELKPHRVTTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2002
|
|
|
QY 61 WEISGPEGKDGAGVYINLSKDTFTKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLN 120
|
|
|
Db 2003 WEISGPEGKDGAGVYINLSKDTFTKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLN 2062
|
|
|
QY 121 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 164
|
|
|
Db 2063 ESHRKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 2106
|
|
|

RESULT 5
US-10-617-320-3169
/ Sequence 3169, Application US/10617320
```

```
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
; THERAPEUTICS
;
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
;
; US-10-617-320-3169
;
; Query Match 99.6%; Score 845; DB 5; Length 637;
; Best Local Similarity 99.4%; Pred. No. 1e-63; Mismatches 0; Indels 0; Gaps 0;
; Matches 163; Conservative 1;
;
; QY 1 TTVKFELMKTGVSSELKPHRTVTIQQKEMSTIVSEEDFILPVYKGELEKGYQFDG 60
; DB 440 TTVKFELMKTGVSSELKPHRTVTIQQKEMSTIVSEEDFILPVYKGELEKGYQFDG 499
;
; QY 61 WEISGFEKGDAGYVNLKDTFIKPVFKIEEKKBEENKPTFDVSKKKDNPQVNHSQLN 120
; DB 500 WEISGFEKGDAGYVNLKDTFIKPVFKIEEKKBEENKPTFDVSKKKDNPQVNHSQLN 559
;
; QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 164
; DB 560 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 603
;
; RESULT 6
; US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
;
; US-09-765-272-68
;
; Query Match 72.5%; Score 615; DB 3; Length 117;
; Best Local Similarity 100.0%; Pred. No. 6.1e-45;
; Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 48 YKGELEKGYQPDGWEISGFEGKGDAGYVNLKDTFIKPVFKIEEKKBEENKPTFDVSK 107
; DB 1 YKGELEKGYQPDGWEISGFEGKGDAGYVNLKDTFIKPVFKIEEKKBEENKPTFDVSK 60
;
; QY 108 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 164
; DB 61 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 117
;
; RESULT 7
; US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

```
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
;
; US-09-765-272-68
;
; Query Match 72.5%; Score 615; DB 3; Length 117;
; Best Local Similarity 100.0%; Pred. No. 6.1e-45;
; Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 48 YKGELEKGYQPDGWEISGFEGKGDAGYVNLKDTFIKPVFKIEEKKBEENKPTFDVSK 107
; DB 1 YKGELEKGYQPDGWEISGFEGKGDAGYVNLKDTFIKPVFKIEEKKBEENKPTFDVSK 60
;
; QY 108 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 164
; DB 61 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 117
;
; RESULT 7
; US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

```
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      72.5%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 6.1e-45;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YGELEKGYQPGWEISGEGKDGAGYVNLKDTPIKPVFKIIEKKEEENKPTFDVSK 107
DB 1 YGELEKGYQPGWEISGEGKDGAGYVNLKDTPIKPVFKIIEKKEEENKPTFDVSK 60

QY 108 KDNQPNVHSQLNESHKEDLQREHSHKSDSTKDVATVTLDKNNISSKTTNNPK 164
DB 61 KDNQPNVHSQLNESHKEDLQREHSHKSDSTKDVATVTLDKNNISSKTTNNPK 117

RESULT 8
US-10-724-972A-5663
; Sequence 5663, Application US/10724972A
; Publication No. US2004014773441
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; PRIOR FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 5663
; LENGTH: 778
; TYPE: PRT
; ORGANISM: S. epidermidis
US-10-724-972A-5663

Query Match      14.0%; Score 119; DB 4; Length 778;
Best Local Similarity 27.0%; Pred. No. 0.17;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSTIVSEDFILPVYK-----GELEKGYQPDGW-- 61
DB 617 EDVNAQSLKP-----ITGNGKQIKQOSVSGTKVLPKSHKVMMLTDTGELTMP-DMTGWTK 671

QY 62 -EISGFE-----GKDDAGYVIN--LSKDTPIKPVFK-----KIEEKKKEENKPTF 107
DB 672 EDVLAPEDLTKIKVSTKNGFPVTNQISKGQIIK-----NKKIEVLSAED 718

QY 108 KDNQPNVHSQLNESHKEDLQREHSHKSDSTKDVATVTLDKNNISSKTTNN 161
DB 719 TDDQEKTDSDSDNKSDDKADHDHNTSSSTKN-----DKSNADSKNDSD 766

RESULT 9
US-10-282-122A-70721
; Sequence 70721, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-23
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70721
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70721

Query Match      13.1%; Score 111.5; DB 4; Length 775;
Best Local Similarity 24.9%; Pred. No. 0.74;
Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSTIVSEDFILPVYK-----GELEKGYQPDGW-- 61
DB 614 EDVNAQSLKP-----ITGNGKQIKQOSVSGTKVLPKSHKVMMLTDTGELTMP-DMTGWTK 668

QY 62 -EISGFE-----GKDDAGYVIN--LSKDTPIKPVFK-----KIEEKKKEENKPTF 103
DB 669 EDVLAPEDLTKIKVSTKNGFPVTNQISKGQIIK-----KIEVLSAEDTDDQEKTDSDS 728

QY 104 DVSKKKNQPNVHSQLNESHKEDLQREHSHKSDSTKDVATVTLDKNN 152
DB 729 DKSKCKOKVDNDSNASSSKNEKSNADSKNDSDSTNETSGS--ERNN 775

RESULT 10
US-10-282-122A-52942
; Sequence 52942, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
```

[illegible]

Db 469 KSEHEARSKAEASSVDYILGWFEGGVPEHKKEENMLSHLYVSSKKENISKENDVDLD 528  
Qy 93 EKKEENKPTFPVSKKONPQVNHSLN-----ESHKEDLQREHHSQKSDS 139  
Db 529 E-KBEAEETEELFEKNEEETSEISDEEBEERKEEENKQKEQENNDQ 587  
Qy 140 TKDVTATVLDKNNISSKSTNN 161  
Db 588 KKDEA-----QNLISKQNNN 604

RESULT 12  
US-10-691-672A-7  
; Sequence 7, Application US/10691672A  
; Publication No. US20050112133A1  
; GENERAL INFORMATION:  
; APPLICANT: DRUIIHE, PIRRE  
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND  
; FILE REFERENCE: 02356.0085  
; CURRENT APPLICATION NUMBER: US/10/691,672A  
; CURRENT FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patent in Ver. 3.3  
; SEQ ID NO 7  
; LENGTH: 188  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
; NAME/KEY: SITE  
; LOCATION: (1)..(188)  
; OTHER INFORMATION: MSP3a to MSP3f  
US-10-691-672A-7

Query Match 12.7%; Score 108; DB 5; Length 188;  
Best Local Similarity 23.2%; Pred. No. 0.26;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
Qy 27 IQNGKMSSTIVSEEDFILPVYKGELEKGYQDFGWEISGF--EGKKGAD-----YVINLS 79  
Db 15 VLKAKEASS-----YDIIL-----GWFFGGVPEHKKEENMLSHLYVSSKD 55  
Qy 80 KDTFIKPVFKKIEEKKE-----ENKPTFDVSKKONPQVNHSLNESHKKE 126  
Db 56 KENISKENDVDLDEKEEAEETEELKEEKEEETESEISEDEEBEERKEEENNDKKK 115  
Qy 127 DLQREHHSQKSDTKDVTATVLDKNNISSKSTNN 161  
Db 116 EQEKEQSNENNDQKDEA-----QNLISKQNNN 145

RESULT 13  
US-10-739-930-6262  
; Sequence 6262, Application US/10739930  
; Publication No. US20040216190A1  
; GENERAL INFORMATION:  
; APPLICANT: Kovalic, David K.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH  
; FILE REFERENCE: 38-21(53377)B  
; CURRENT APPLICATION NUMBER: US/10/739,930  
; CURRENT FILING DATE: 2003-12-18  
; NUMBER OF SEQ ID NOS: 11088  
; SEQ ID NO 6262  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C271270\_1.p  
US-10-739-930-6262

Query Match 12.7%; Score 107.5; DB 5; Length 470;  
Best Local Similarity 20.1%; Pred. No. 0.88;

Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;  
Qy 19 KPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDFGWEISGFE-----GKK 70  
Db 82 RENRVTDVTVQNNNGESK-----YVQDLARRIRYDE-EATQSAQRIDHPNQK 129  
Qy 71 DAGYVINLSKDTFIKPVFKKIEEKKEENKPTFDVSKKDN----- 111  
Db 130 NVGITEKAPENSPIEETSHRVDDNKRINNQNFTAAKSENNAVSRVSFGADHKRAEYMGK 189  
Qy 112 PQVNHSQLNE-----SHRKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTNNPNK 164  
Db 190 PMENRDQVRQTESAEKSHRKENVTK3EKPQDEGVKKTAKDKRNRKKEEKTESINK 248

RESULT 14  
US-09-820-843A-107  
; Sequence 107, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 107  
; LENGTH: 665  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: hypothetical protein  
; NAME/KEY: misc feature  
; OTHER INFORMATION: gi|3845248  
US-09-820-843A-107

Query Match 12.5%; Score 106; DB 3; Length 665;  
Best Local Similarity 24.3%; Pred. No. 1.8;  
Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;  
Qy 10 KDTGEVSELKPHRYT-VTIQNGKMSSTIVSEEDFILPVYKGELEKGYQDFGWEI--SGF 66  
Db 127 EERKNKNSDLHRQNELNLSGK-----NEQDI-----NKNEKGKQ-----DISNSNA 169  
Qy 67 EGKKGAGYVINLSKDTFIKPVFKKIEEKKE-----EENKPTFD-----VSKKKDNPN 112  
Db 170 ENKGD-----VKEGVKLEBKKEEKISDDHKVEENKKSDDHKVEENKKSDDH 217  
Qy 113 QVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTNNPNK 164  
Db 218 KVEENKKSDDHKIEVKKVEEHEDEEBE-----DKKEKSENKKNKDNK 261

RESULT 15  
US-10-732-923-22820  
; Sequence 22820, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 22820  
; LENGTH: 540  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana

GENERAL INFORMATION:  
 APPLICANT: Wang, Liangou  
 APPLICANT: Zamudio, Carlos  
 APPLICANT: Malone, Cheryl  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsein, Kari  
 APPLICANT: Zyskind, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carr, Grant

; PRIOR APPLICATION NUMBER: 60/157,832  
 ; PRIOR FILING DATE: 1999-10-05  
 ; PRIOR APPLICATION NUMBER: 60/160,191

QY 137 -----SDST-----KDVATVLDK--NNIS-SKSTNNPN 163  
 DB 726 GNIMDLITSTNDNKNKNDHSSQVIONVSCITRDKEGDNKINTHTNNPN 779

RESULT 20  
 US-10-732-923-18783  
 ; Sequence 18783, Application US/10732923  
 ; Publication No. US20050108791A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edgerton, Michael D  
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
 ; FILE REFERENCE: 38-15(52796)C  
 ; CURRENT APPLICATION NUMBER: US/10/732,923  
 ; CURRENT FILING DATE: 2003-12-10  
 ; PRIOR APPLICATION NUMBER: 10/310,154  
 ; PRIOR FILING DATE: 2002-12-04  
 ; NUMBER OF SEQ ID NOS: 24149  
 ; SEQ ID NO 18783  
 ; LENGTH: 973  
 ; TYPE: PRT  
 ; ORGANISM: Plasmodium yoelii yoelii  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)..(973)  
 ; OTHER INFORMATION: unsure at all Xaa locations  
 US-10-732-923-18783

Query Match 11.9%; Score 101; DB 5; Length 973;  
 Best Local Similarity 21.0%; Pred. No. 7.8; Indels 12; Gaps 4;  
 Matches 34; Conservative 43; Mismatches 73; Indels 12; Gaps 4;  
 QY 7 ILNKDTGB-----VSELKPHRVTVTIQNGKEMSSSTIVSBEEDFILPVYKGELEKGYQFDG 60  
 DB 581 LVNNTNQNQNSVDKLETTYNINKNDQKGCNSTYKT-----LIQHSKSGSKGNFKFIH 636

QY 61 WEISGFEKGDAGYVYNLSKDTFKIPVKKIEEKEEENKPTFDVSKKDNPNVHSQLN 120  
 DB 637 TKINNISNRHARGYVSLSKDNNEVRNDNIEVRNDNIEVRNDNIEVRNDNIEVRNDNIN 696

QY 121 ESHRKEDL-QREEHSQKSDSTKDVATVLDKNNISSKSTNN 161  
 DB 697 DQNEKNISNETSPNSKSEXT-FVTATSNKSKODNINKNSND 737

RESULT 21  
 US-10-732-923-16976  
 ; Sequence 16976, Application US/10732923  
 ; Publication No. US20050108791A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Edgerton, Michael D  
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
 ; FILE REFERENCE: 38-15(52796)C  
 ; CURRENT APPLICATION NUMBER: US/10/732,923  
 ; CURRENT FILING DATE: 2003-12-10  
 ; PRIOR APPLICATION NUMBER: 10/310,154  
 ; PRIOR FILING DATE: 2002-12-04  
 ; NUMBER OF SEQ ID NOS: 24149  
 ; SEQ ID NO 16976  
 ; LENGTH: 1373  
 ; TYPE: PRT  
 ; ORGANISM: Plasmodium falciparum 3D7  
 US-10-732-923-16976

Query Match 11.9%; Score 100.5; DB 5; Length 1373;  
 Best Local Similarity 21.7%; Pred. No. 13; Indels 57; Gaps 9;  
 Matches 40; Conservative 33; Mismatches 54; Indels 57; Gaps 9;  
 QY 6 FILNKDTGEVSP-----LKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQFDGW 61  
 DB 418 YVNSDRGVMEESIIMPLPHHVQILL-----LSATV-----PNT-----LEFADW 457

QY 62 -----EISGPEGKK-----DAGVINLSKDTFKIPVKF-----KIEEKEEE 98



```
Db 458 VQFTQKQVISTKRPVPLHLYIVYDVSVLVMDKRNKFFYSAPKEIYVIRKQBQAN 517
Qy 99 NKPTFDVSKKONPOVNHSQLNESH--RKEDLQREHSHSQSDSTKDVATVLDKNNISSKS 157
Db 518 NKNTQITSGSNNTSSNLKKNYDKNKYLITNNKENDNT-----QNNNNNN 568
Qy 158 ITNN 161
Db 569 NNNN 572

RESULT 22
US-10-732-923-22588
; Sequence 22588, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; PRIOR FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22588
; LENGTH: 3127
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-22588

Query Match 11.8%; Score 100; DB 5; Length 3127;
Best Local Similarity 24.2%; Pred. No. 40;
Matches 52; Conservative 34; Mismatches 71; Indels 58; Gaps 9;

Qy 4 KEPILNDTGVSELSKPHRVTVTIQNG-----KMSSTIVSEEDFILPVYKGELEKGYQ 57
Db 1969 KKF--KRNTSYVLESPLHLIGDIVDNNIKRKKKKKEIKTIVSDDMFTSPVNIKEYNYEQ 2026
Qy 58 PDGWEISG---PEGKKDAGYVNLSDTFFIKPVFKKIEKKKEENK-----100
Db 2027 ERKKEIVGNLSYDKTKLFPFFIKFTKEGRKK--KKIEKKKEKKENNNFLYNDYSS 2084
Qy 101 ---PTF-----DVSKKONPOVNHSQL-----NESHK---EDLQR 130
Db 2085 YSSPKYGDNNFNVIKIRERKDFQKCDHPNPNFSEKFLNYPNPKNKNKNNKNVRR 2144
Qy 131 EHSQKSDTKD-VTATVLDKNNISSKSTNNPNK 164
Db 2145 NEYPNTSSKDGVSYNFLSDSLFSSDNEYSSDNE 2179

RESULT 23
US-10-473-576-22
; Sequence 22, Application US/10473576
; Publication No. US20040101884A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: LU, DYUNG AINA M.
; APPLICANT: ARVIZU, CHANDRA S.
; APPLICANT: GANDHI, AMEENA R.
; APPLICANT: HAFALIA, APRIL J.A.
; APPLICANT: DING, LI
; APPLICANT: LU, YAN
; APPLICANT: RAMKUMAR, JAYALAXMI
; APPLICANT: SWARNAKAR, ANITA
; APPLICANT: TANG, Y. TOM
; APPLICANT: YUE, HENRY
; APPLICANT: TRAN, BAO
; APPLICANT: LEE, SOO YUEN
; APPLICANT: WARREN, BRIDGET A.
; APPLICANT: NGUYEN, DANNIEL B.
; APPLICANT: THANGAVELU, KAVITHA
```

```
; APPLICANT: YAO, MONIQUE G.
; APPLICANT: ELLIOTT, VICKI S.
; APPLICANT: BAUGHN, MARIAH R.
; APPLICANT: EMERLING, BROOKE M.
; APPLICANT: LAL, PREETI G.
; APPLICANT: GIETZEN, KIMBERLY J.
; APPLICANT: BECHA, SHANYA D.
; APPLICANT: MARQUIS, JOSEPH P.
; APPLICANT: KABLE, AMY E.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PP-0921 USN
; CURRENT APPLICATION NUMBER: US/10/473,576
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US02/09809
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: US 60/280,387
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 60/282,335
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: US 60/286,663
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/285,484
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/350,702
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/351,749
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 1384
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 7506096CD1
US-10-473-576-22

Query Match 11.7%; Score 99.5; DB 4; Length 1384;
Best Local Similarity 21.7%; Pred. No. 16;
Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;

Qy 4 KEPILNDTGVSELSKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQF-----58
Db 1013 KEFTMLQNEQISQLK-KBIERTQQRKEMESVMKQBQYIATQYKEAIDLQELRLTRE 1071
Qy 59 -----DGWEISGPEKKDA-----GYVINLSK-----80
Db 1072 QVQNSHTELAEARHQVQAQREIERLSSELEDMKQLSKEDAHGNHNLABELGASKVREAH 1131
Qy 81 -----DTFIPKVPFKKIEKK-----BENKPTDVSKKONPOVNHSQLNESHK 126
Db 1132 LEARMQAIEKKLSAEVESLSKEAYHMEMISHQENHAKWKIS--ADSKSSVQQLNEQLEKA 1189
Qy 127 DLQREHSHSQSDSTKDVTVATVLDKNNI 153
Db 1190 KLELEE---AQDTVSNLHQVQDRNEV 1213

RESULT 24
US-10-473-576-2
; Sequence 2, Application US/10473576
; Publication No. US20040101884A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: LU, DYUNG AINA M.
; APPLICANT: ARVIZU, CHANDRA S.
; APPLICANT: GANDHI, AMEENA R.
; APPLICANT: HAFALIA, APRIL J.A.
; APPLICANT: DING, LI
; APPLICANT: LU, YAN
; APPLICANT: RAMKUMAR, JAYALAXMI
; APPLICANT: SWARNAKAR, ANITA
```

APPLICANT: TANG, Y. TOM  
APPLICANT: YUB, HENRY  
APPLICANT: TRAN, BAO  
APPLICANT: LEE, SOO YUEN  
APPLICANT: WARREN, BRIDGET A.  
APPLICANT: NGUYEN, DANNIEL B.  
APPLICANT: THANGAVELU, KAVITHA  
APPLICANT: YAO, MONIQUE G.  
APPLICANT: ELLIOTT, VICKI S.  
APPLICANT: BAUGHN, MARIAH R.  
APPLICANT: EMERLING, BROOKE M.  
APPLICANT: LAL, PREETI G.  
APPLICANT: GIETZEN, KIMBERLY J.  
APPLICANT: BECHA, SHANYA D.  
APPLICANT: MARQUIS, JOSEPH P.  
APPLICANT: KABLE, AMY E.  
TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT  
FILE REFERENCE: PF-0921 USN  
CURRENT APPLICATION NUMBER: US/10/473,576  
PRIOR FILING DATE: 2003-09-29  
PRIOR APPLICATION NUMBER: PCT/US02/09809  
PRIOR FILING DATE: 2002-03-29  
PRIOR APPLICATION NUMBER: US 60/280,387  
PRIOR FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: US 60/282,335  
PRIOR FILING DATE: 2001-04-05  
PRIOR APPLICATION NUMBER: US 60/286,663  
PRIOR FILING DATE: 2001-04-13  
PRIOR APPLICATION NUMBER: US 60/285,484  
PRIOR FILING DATE: 2001-04-19  
PRIOR APPLICATION NUMBER: US 60/350,702  
PRIOR FILING DATE: 2002-01-18  
PRIOR APPLICATION NUMBER: US 60/351,749  
PRIOR FILING DATE: 2002-01-25  
NUMBER OF SEQ ID NOS: 46  
SOFTWARE: PERL Program  
SEQ ID NO 2  
LENGTH: 1404  
TYPE: PRT  
ORGANISM: Homo sapiens  
NAME/KEY: misc feature  
FEATURE:  
OTHER INFORMATION: Incyte ID No: 3125036CD1  
US-10-473-576-2

Query Match 11.7%; Score 99.5; DB 4; Length 1404;  
Best Local Similarity 21.7%; Pred. No. 16;  
Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;  
Qy 4 KEFILNKDGEVSELKPHRVTTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQF----- 58  
Db 1033 KEFILNLQNEQISQLK-KEIERTQQRKMEVSMKEQEQTATQYKEAIDLGQELRLTRE 1091  
Qy 59 -----DGEIISGPEGKKDA-----GVINLSK----- 80  
Db 1092 QVQNSHTELAARHQVQQRERIERLSSELEDMKQLSKEKDAHGNHAEELGASKVREAH 1151  
Qy 81 -----DTFTKPVFKIEEKE-----BENKPTFDVSKKKQNPQVNHLSHSHAKE 126  
Db 1152 LEARMQABTKGLSAEVESLEKAYHMEMISHQENHAKWKIS--ADSKGSVQQLNQLEKA 1209  
Qy 127 DLQREHSQKSDTKDVTATVLDKNNI 153  
Db 1210 KLELEB---AQDTVSNLHQVQQRNEV 1233

## RESULT 25

US-10-282-122A-53254  
Sequence 53254, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 53254  
LENGTH: 1184  
TYPE: PRT  
ORGANISM: Clostridium difficile  
US-10-282-122A-53254

Query Match 11.5%; Score 97.5; DB 4; Length 1184;  
Best Local Similarity 26.3%; Pred. No. 20;  
Matches 46; Conservative 30; Mismatches 70; Indels 29; Gaps 9;  
Qy 14 EVSELKPHRVTTIQ-NGKEMSSSTIV--SEEDF--ILPVYKGELEKGYQFDGWEISGPEG 68  
Db 240 ELSEVNEHRKVIEKELNEKEQKVVEKQEDINKEVEVLQDVIEKSVDIYN-SIKGVIS 298  
Qy 69 KQDAGVYNLSKDTF-----IKPVFKIEEKE-----BENKPTFDVSKKKD 110  
Db 299 KKE5--QINLTKERIRNFTNEISRNLKIDKEKLANENKQYIKELESNK--LSGSEBLS 354  
Qy 111 NPQVNHLSNESHKEDLQREHSCSDSTKDVATVLD-KNNISSKSTTNNPNK 164  
Db 355 TLQENIKVLESGKQKQIKLESNLNELLAKESIIDLKQKQFNSKLTSLNANK 409

RESULT 26  
US-10-755-889-615  
Sequence 615, Application US/10755889  
Publication No. US20040171823A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB  
FILE REFERENCE: D0284 NP  
CURRENT APPLICATION NUMBER: US/10/755,889  
CURRENT FILING DATE: 2004-01-13  
PRIOR APPLICATION NUMBER: U.S. 60/440,068  
PRIOR FILING DATE: 2003-01-14

; PRIOR APPLICATION NUMBER: U.S. 60/469,757  
; PRIOR FILING DATE: 2003-05-12  
; NUMBER OF SEQ ID NOS: 823  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 615  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-755-889-615

Query Match 11.5%; Score 97.5; DB 4; Length 2468;  
Best Local Similarity 24.8%; Pred. No. 49;  
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKTGVSSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDFGWBI 63  
DB 584 EKVMVKDKPKVKTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623

QY 64 SGFEGKDGAGYVNLSDTFIKPVFKIEKKEENKPTFDVSKKNDPQVNHSQLNESH 123  
DB 624 QATDVKPKAAKEKTVKKTVP-----EDKKEEKEPKKGVAKKEDKTP-----KKEBKP 675

QY 124 RKEDLQRE-----EHSOKSDSTKDV 143  
DB 676 KKEEVKKEVKKEIKKEEKKPKKEV 700

RESULT 27  
US-10-489-740-216  
; Sequence 216, Application US/10489740  
; Publication No. US20050112574A1  
; GENERAL INFORMATION:  
; APPLICANT: Biomedics Limited  
; TITLE OF INVENTION: P9  
; FILE REFERENCE: Angiogenesis PCT  
; CURRENT APPLICATION NUMBER: US/10/489,740  
; CURRENT FILING DATE: 2004-03-15  
; NUMBER OF SEQ ID NOS: 216  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 216  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-489-740-216

Query Match 11.5%; Score 97.5; DB 5; Length 2468;  
Best Local Similarity 24.8%; Pred. No. 49;  
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKTGVSSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDFGWBI 63  
DB 584 EKVMVKDKPKVKTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623

QY 64 SGFEGKDGAGYVNLSDTFIKPVFKIEKKEENKPTFDVSKKNDPQVNHSQLNESH 123  
DB 624 QATDVKPKAAKEKTVKKTVP-----EDKKEEKEPKKGVAKKEDKTP-----KKEBKP 675

QY 124 RKEDLQRE-----EHSOKSDSTKDV 143  
DB 676 KKEEVKKEVKKEIKKEEKKPKKEV 700

RESULT 28  
US-10-450-763-46995  
; Sequence 46995, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790CIP3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763  
; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631

; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 46995  
; LENGTH: 2519  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (1040)..(1091)  
; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,  
; OTHER INFORMATION: accession number BL00412D, p-value=2.432e-11, raw score of 16.5  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (1919)..(2122)  
; OTHER INFORMATION: Neuraxin and MAP1B proteins domain identified by Pfam,  
; OTHER INFORMATION: accession name MAP1B\_neuraxin, E-value=1.9e-59, Pfam score of 1  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)....(2519)  
; OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
US-10-450-763-46995

Query Match 11.5%; Score 97.5; DB 5; Length 2519;  
Best Local Similarity 24.8%; Pred. No. 50;  
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKTGVSSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQDFGWBI 63  
DB 635 EKVMVKDKPKVKTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 674

QY 64 SGFEGKDGAGYVNLSDTFIKPVFKIEKKEENKPTFDVSKKNDPQVNHSQLNESH 123  
DB 675 QATDVKPKAAKEKTVKKTVP-----EDKKEEKEPKKGVAKKEDKTP-----KKEBKP 726

QY 124 RKEDLQRE-----EHSOKSDSTKDV 143  
DB 727 KKEEVKKEVKKEIKKEEKKPKKEV 751

RESULT 29  
US-10-425-115-205148  
; Sequence 205148, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 205148  
; LENGTH: 898  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MFT4577\_118684C.1.pep  
US-10-425-115-205148

Query Match 11.4%; Score 96.5; DB 4; Length 898;  
Best Local Similarity 24.1%; Pred. No. 17;  
Matches 42; Conservative 34; Mismatches 73; Indels 25; Gaps 7;  
QY 7 ILNKDTGVSSELKPHRVTVTIQNGKEMSTIVSEEDP-----ILPVYKGELEK-----54

Db 557 ILLKMTESVVA--AASTATEKGKEQAEILEEBDFEQDLLGQLTDAKAEULKCAI 614  
Qy 55 --GYQFDGWEISGF-EGKXDAGYVINLSKDTFIKVFVKIEBKEEENKPTP-DVSKKKD 110  
Db 615 ACGYKPGATLFGVNEGK-----LRSFRQSILLSKVLKMQQDLBEEKNAVIQDLAEKVE 669  
Qy 111 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164  
Db 670 NYEASLKKXKPTIQLLEIMVKEHGALEKDFVIQTM--EGSLAEVQTEENKLG 721

RESULT 30  
US-10-381-596A-2  
; Sequence 2, Application US/10381596A  
; Publication No. US20040014178A1  
; GENERAL INFORMATION:  
; APPLICANT: Biostapro AB  
; TITLE OF INVENTION: von Willebrand factor-binding proteins from  
; TITLE OF INVENTION: Staphylococci  
; FILE REFERENCE: 110059600  
; CURRENT APPLICATION NUMBER: US/10/381,596A  
; CURRENT FILING DATE: 2003-07-02  
; PRIOR APPLICATION NUMBER: SE 0003573-3  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2060  
; TYPE: PRT  
; ORGANISM: Staphylococcus lugdunensis  
US-10-381-596A-2

Query Match 11.4%; Score 96.5; DB 4; Length 2060;  
Best Local Similarity 25.7%; Pred. No. 48;  
Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;  
Qy 14 EYSELKPHRVTVTIO-NGKEMSTIVSEBDFILPVYKGELEKGYQFDG--WEISGFEGK 70  
Db 1914 DIDEUK---ITIVDTNGRE-----IVPSRKQLPP-EQFIQDQWQYGHK--- 1955  
Qy 71 DAGYVINLSKDTFIKVPFKIEE-----KKEENKPTDFVSKKKNPQVNHSQLNESH 123  
Db 1956 -----IEKDGITTVYKKVENAVPAKQKTKEN--TQSSQPKHTPPQVKQQLVKYHN 2006  
Qy 124 RKE--DLQREHSQKSDSTKDVATVLDKNNI 153  
Db 2007 VKEQRSIEKSEHTDMHVSLSLPETGETANKGL 2038

RESULT 31  
US-10-732-923-4285  
; Sequence 4285, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 4285  
; LENGTH: 1419  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-10-732-923-4285

Query Match 11.3%; Score 96; DB 5; Length 1419;  
Best Local Similarity 21.9%; Pred. No. 33;  
Matches 49; Conservative 36; Mismatches 69; Indels 70; Gaps 11;  
Qy 6 FILNKDTGVSSEL-----KPHRVTVTIQNGKEMSTIVSEBDFILPVI-----KG 50

Db 1031 YIINLGGYPEELEYDQNAKHKDKHYCTLSIQEIKKUIQKFLNEETFLTKYGYSPFKI 1090  
Qy 51 ELEKGYQFDGWEISGFEGKGD-----AGYVINLSKDTFIKVPFKK-----IEEK 94  
Db 1091 SLAINMSIDHY-----FSHMKDNLRVICEPGRYVMVAASSTLAVKIIGKRRTPTFOGIMLKEL 1146  
Qy 95 KEEENKPTP--DVSKKKDNQVNH---SQLNESHKEDLQREHSQK----- 136  
Db 1147 KQHYDPLNFAQENKQDETKINENNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDLIIT 1206  
Qy 137 --SDST-----KQVTVATVLDK--NNIS-SKSTTNNPN 163  
Db 1207 STNDSTKKNHSSSQVIVQVSCITIRKDEGDNINIKINTINNPN 1250

RESULT 32  
US-10-732-923-8760  
; Sequence 8760, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 8760  
; LENGTH: 1350  
; TYPE: PRT  
; ORGANISM: Plasmodium yoelii yoelii  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1350)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-732-923-8760

Query Match 11.3%; Score 95.5; DB 5; Length 1350;  
Best Local Similarity 23.2%; Pred. No. 34;  
Matches 45; Conservative 35; Mismatches 71; Indels 43; Gaps 10;  
Qy 1 TTVKEFILNKDTGVSSELKPHRVTVTIQNGK-----EMSSTIVSEBDFILPVYKGELEK 54  
Db 852 STIKKFLKNN-----KGMIFLDLGERSKKWKVWNTACTTKTKKAILYGHKAEARG 903  
Qy 55 GYQFDG-----EISGFEGKXDAGYVINLSKDTFIKVPFKKIEBKEE---NKPTPDV 105  
Db 904 GHDFQFNNKLEDEKIEBKWN-NYHINQK---IKEIISQTEKEDFEKIIINIEFLT 959  
Qy 106 SKKKD-----NPQV-NHSQLNESHKEDLQRE-----EHSQKSDSTKDVATVLDK 150  
Db 960 HHAKNIYNLINLPNINLDSNIGDSIEEKDVSKDIQEDGNEGKYGKSGFNGI-VNIYSK 1018  
Qy 151 NNISKSTNNPNK 164  
Db 1019 SNVGSSETRNDSK 1032

RESULT 33  
US-10-732-923-4235  
; Sequence 4235, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149

```
; SEQ ID NO 4235
; LENGTH: 1434
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-4235

Query Match
Best Local Similarity 11.3%; Score 95.5; DB 5; Length 1434;
Matches 52; Conservative 31; Mismatches 70; Indels 75; Gaps 12;

QY 6 FILNKDTGEVSEL-----KPHRVTVTIQNGKMSSTIVSEEDFILPVY-----KG 50
Db 1043 YIINLGGYPELEVDNAKHKDIHYCTLSLQIKDKQKPLNEETFLTKYGYSPKI 1102

QY 51 ELEKGYQDFGWISGPEGKKO-----AGVINLSKDTFTK-----PVFKKI----- 91
Db 1103 SLAINKSIDHY-----FSHKMCLRVICBPGRTVAASSTLAVKIIGKRPTFGIMLKD 1158

QY 92 -----BEKKEBENKTFP--DVSKKDNQVNHSQLNESHKKE-----DL 128
Db 1159 KDHYDPLNPAQOENKKQDETINHNNDNNDNNDNNDNNDNNDNNDNNDNNDNNDN 1218

QY 129 -----OREEHSQKSDSKQVATVLDK--NNIS-SKSTNNPN 163
Db 1219 IITSTNDSTNKNDHS--SSQVIONVSTIRDEGDKINHTINPN 1265

RESULT 34
US-10-732-923-22709
; Sequence 22709, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22709
; LENGTH: 3124
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-22709

Query Match
Best Local Similarity 11.1%; Score 94.5; DB 5; Length 3124;
Matches 50; Conservative 33; Mismatches 74; Indels 55; Gaps 8;

QY 4 KEFILNKDTGEVSELKPHRVTVTIQNG-----KMSSTIVSEEDFILPVYKGELEKGYQ 57
Db 1969 KKF--KRNTSVYLESPLHLIGDIVDNNIKRKKKKKEIKTIVSDDMFTSPVNIKEYNNEQ 2026

QY 58 FDGWEISG---PEGKKDAGVINLSKDTFTKPVFKKIEKKEBENKTFP----- 103
Db 2027 ERKKEIVGNLSYDKTKICPFKFTKEGRIKK--NKIEKKEKKEYNFNFLYNDYSSYSS 2084

QY 104 -----DVSKKDNQVNHSQL-----NESHKKE---DLQREH 133
Db 2085 PRYGDNNFNVIKYIRERKDFQKFDHPNFNFKPLHNNYPMKNQKKNVNRNEY 2144

QY 134 SQKSDSTKD-VTATVLDKNNISKSTNNPNK 164
Db 2145 PNYTSSKDGVSYNFLSDLSFSSDNEYSSDNE 2176

RESULT 35
US-10-282-122A-70294
; Sequence 70294, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangou

; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvakind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70294
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-282-122A-70294

Query Match
Best Local Similarity 11.0%; Score 93.5; DB 4; Length 645;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 1 TTVEFILNKDTGR-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGBLB-KGY 56
Db 327 SAITFQNVQPTNEKMTDLQDTKYVYVESVENNESMMDTFVKH-----PIKTGMLNGKKY 381

QY 57 QP-----DGWEISGPEGKK-----D 71
Db 382 MVMETTNDNDYKQDPVGEQVRVTISKDAKNKTRTIIPYVEGKTLDAIVKRVHVKTIIDYD 441

QY 72 AGVYVNL-SKDTFTKPVFKKIEKKEBENKTFDV-----SKKDNQVNHSQLNESHK 125
Db 442 GQHVRIVDKEAFTKANTDKSNKEQQDNNAKKEATATPSKPTSPVEKESQKDSQKD 501

QY 126 EDLQ-----REEHSQKSDTKDVT-ATVLDKNNISKSTNNPNK 164
Db 502 DNKQLPSVEKENDASSESGKDKTPTKPTKGEVSSSTI--PTK 543

RESULT 36
US-10-470-048B-414
; Sequence 414, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; ANTIGENS TO A SPECIFIC PATHOGEN
```

```
; FILE REFERENCE: SONN:035US
; CURRENT APPLICATION NUMBER: US/10/470,048B
; CURRENT FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 414
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-470-048B-414

Query Match      11.0%; Score 93.5; DB 5; Length 645;
Best Local Similarity 21.9%; Pred. No. 21;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 1 TTVKEFILNKDTGGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 327 SAITEFQNVQPTNEKMTDLQDTKYVYVESVENNESMMDTFVKH-----PIKTGMLNGKKY 381
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 57 QF-----DGWEISGPEGKK-----D 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 382 MVMETTTNDYDKDFWVEGQVRVTSKDAKNTRTIIFPVVEGKTYLDAIVKVHVKTIDYD 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 AGVYINL-SKDTFIKPVFKKIEEKKSEENKPTFDV-----SKKDNQPNVHSQLNESHRK 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 GOYHVRIVDKEAFTKANTDKSNKKEQDNSAKKEATPATPSKPTSPVPEKESQKDSQKD 501
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 EDLQ-----REHSQKSDSTKQVT-ATVLDKNNISKSSTNNPNK 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 502 DNKQLPSVEKENDASSESGDKTPTATKPTKGEVSSSTT--PTK 543
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 37
US-10-172-502-10
; Sequence 10, Application US/10172502
; Publication No. US2003018583A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-10

Query Match      11.0%; Score 93.5; DB 4; Length 654;
Best Local Similarity 21.9%; Pred. No. 21;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 1 TTVKEFILNKDTGGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 SAITEFQNVQPTNEKMTDLQDTKYVYVESVENNESMMDTFVKH-----PIKTGMLNGKKY 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 57 QF-----DGWEISGPEGKK-----D 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 MVMETTTNDYDKDFWVEGQVRVTSKDAKNTRTIIFPVVEGKTYLDAIVKVHVKTIDYD 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 AGVYINL-SKDTFIKPVFKKIEEKKSEENKPTFDV-----SKKDNQPNVHSQLNESHRK 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 GOYHVRIVDKEAFTKANTDKSNKKEQDNSAKKEATPATPSKPTSPVPEKESQKDSQKD 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 EDLQ-----REHSQKSDSTKQVT-ATVLDKNNISKSSTNNPNK 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 DNKQLPSVEKENDASSESGDKTPTATKPTKGEVSSSTT--PTK 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 38
US-11-020-509-10
; Sequence 10, Application US/11020509
; Publication No. US20050106648A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .
; FILE REFERENCE: P07263US02/BAS
; CURRENT APPLICATION NUMBER: US/11/020,509
; CURRENT FILING DATE: 2004-12-27
; PRIOR APPLICATION NUMBER: US 10/172,502
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-11-020-509-10

Query Match      11.0%; Score 93.5; DB 6; Length 654;
Best Local Similarity 21.9%; Pred. No. 21;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 1 TTVKEFILNKDTGGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 SAITEFQNVQPTNEKMTDLQDTKYVYVESVENNESMMDTFVKH-----PIKTGMLNGKKY 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 57 QF-----DGWEISGPEGKK-----D 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 MVMETTTNDYDKDFWVEGQVRVTSKDAKNTRTIIFPVVEGKTYLDAIVKVHVKTIDYD 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 72 AGVYINL-SKDTFIKPVFKKIEEKKSEENKPTFDV-----SKKDNQPNVHSQLNESHRK 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 GOYHVRIVDKEAFTKANTDKSNKKEQDNSAKKEATPATPSKPTSPVPEKESQKDSQKD 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 EDLQ-----REHSQKSDSTKQVT-ATVLDKNNISKSSTNNPNK 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 DNKQLPSVEKENDASSESGDKTPTATKPTKGEVSSSTT--PTK 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
RESULT 39
US-10-282-122A-52510
; Sequence 52510, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
```

Search completed: April 24, 2006, 15:43:36  
Job time : 83.1834 secs

; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 52510  
; LENGTH: 932  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-10-282-122A-52510

Query Match 11.0%; Score 93.5; DB 4; Length 932;  
Best Local Similarity 24.4%; Pred. No. 32;  
Matches 42; Conservative 30; Mismatches 53; Indels 47; Gaps 9;  
QY 8 LNKDGEVSELKPHRTVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFE 67  
DB 735 LTKLEDNIEGVNYSVS-----ELENEIV----FLRKIKGGADQSY-----GIE 777  
QY 68 GKGDAGY---VINLSKDTIKVPFKIEKKEE-----ENKPTFDVSKKKN-- 111  
DB 778 VAKLAGLPSFVINRAKE-----ILQHIEGDKKEENSINTAPSKYKDYIEVSKDTSNTK 832  
QY 112 ----PQVNHSQLNESHRK---EDLQREHSQKSDSTKQVATVLDKNMISSK 156  
DB 833 NNLSGEIKDITLSETNTDTIIEDESTKEH--LSSNKKQINCRIINDEKSIKKE 882

RESULT 40  
US-10-437-963-122282  
; Sequence 122282, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 122282  
; LENGTH: 869  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_25224C.1.pap  
US-10-437-963-122282

Query Match 11.0%; Score 93; DB 4; Length 869;  
Best Local Similarity 45.1%; Pred. No. 33;  
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;  
QY 104 DVSKKKDNPPQVNHSQLNESHRKEDLQREHSQKSDSTKQVTA--TVLDKN 152  
DB 617 DASKKDNHQSEGNL--SHRDEDPTRKKKQKTNATSDACAQEVVTEKN 665

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 25.4989 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773  
Perfect score: 848  
Sequence: 1 TTVEFILNKDTGEVSELKP.....ATVLDKNKISSKSTNNPNK 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgm2\_6/ptodata/1/iaa/5 COMB.pep:\*  
2: /cgm2\_6/ptodata/1/iaa/6 COMB.pep:\*  
3: /cgm2\_6/ptodata/1/iaa/H COMB.pep:\*  
4: /cgm2\_6/ptodata/1/iaa/PTUS COMB.pep:\*  
5: /cgm2\_6/ptodata/1/iaa/RE COMB.pep:\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles!.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	848	100.0	773	2	US-09-590-991-8
2	845	99.6	637	2	US-09-107-433-3169
3	845	99.6	2138	2	US-09-583-110-5274
4	615	72.5	117	2	US-08-961-083-68
5	615	72.5	117	2	US-09-536-784-68
6	615	72.5	117	2	US-09-765-271-68
7	615	72.5	117	2	US-09-765-272A-68
8	119	14.0	746	2	US-09-710-279-652
9	119	14.0	778	2	US-09-134-001C-3868
10	101.5	12.0	347	2	US-09-248-796A-16224
11	97.5	11.5	2468	2	US-09-976-594-726
12	97.5	11.5	2468	2	US-09-538-092-1135
13	97.5	11.5	2522	2	US-09-949-016-10237
14	95	11.2	348	2	US-09-538-092-1316
15	93.5	11.0	654	2	US-10-172-502-10
16	93	11.0	280	2	US-09-248-796A-17646
17	92	10.8	299	2	US-09-710-279-1888
18	92	10.8	309	2	US-09-134-001C-5667
19	91	10.7	442	2	US-09-134-001C-3033
20	91	10.7	902	2	US-09-134-001C-5157
21	90.5	10.7	743	2	US-08-910-925-3
22	90.5	10.7	743	2	US-09-949-016-6261
23	90.5	10.7	758	2	US-09-949-016-8288
24	89.5	10.6	262	2	US-09-248-796A-21451
25	89.5	10.6	402	2	US-09-464-483-4
26	89.5	10.6	402	2	US-09-414-664-4
27	89.5	10.6	511	2	US-09-198-452A-509

28	89.5	10.6	511	2	US-09-438-185A-475	Sequence 475, Appl
29	89.5	10.6	529	2	US-09-464-483-2	Sequence 2, Appli
30	89.5	10.6	529	2	US-09-414-664-2	Sequence 2, Appli
31	89.5	10.6	553	2	US-09-248-796A-16588	Sequence 16588, A
32	87.5	10.3	825	2	US-09-248-796A-16538	Sequence 16538, A
33	87.5	10.3	1315	2	US-09-200-650E-5	Sequence 5, Appli
34	87.5	10.3	1702	2	US-08-296-791-5	Sequence 5, Appli
35	87.5	10.3	1702	2	US-09-839-996-5	Sequence 5, Appli
36	87.5	10.3	1702	2	US-10-080-505-5	Sequence 5, Appli
37	87.5	10.3	1702	2	US-10-645-655-5	Sequence 5, Appli
38	87.5	10.3	1702	4	PCT-US95-10661A-5	Sequence 5, Appli
39	86.5	10.2	277	2	US-09-830-230A-651	Sequence 651, App
40	86.5	10.2	732	1	US-08-533-669A-18	Sequence 18, Appl
41	86.5	10.2	732	2	US-09-183-861-18	Sequence 18, Appl
42	86.5	10.2	732	2	US-09-022-765-18	Sequence 18, Appl
43	86.5	10.2	732	2	US-09-551-974A-18	Sequence 18, Appl
44	86.5	10.2	732	2	US-09-565-501A-18	Sequence 18, Appl
45	86.5	10.2	732	2	US-09-639-206A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1  
US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590.991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%	Score 848;	DB 2;	Length 773;
Best Local Similarity	100.0%	Pred. No. 6.2e-79;		
Matches 164;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TTTVEFILNKDTGEVSELKPHRVTTTQNGKMSSTIVSEEDFILPVYKGLEKGYQFDG	60	
Db	610	TTTVEFILNKDTGEVSELKPHRVTTTQNGKMSSTIVSEEDFILPVYKGLEKGYQFDG	669	
Qy	61	WEISGFEGKKDAGVYINLSKDTFKPVFKLEKKKEBKPTFDVSKKKNPQVNHSQLN	120	
Db	670	WEISGFEGKKDAGVYINLSKDTFKPVFKLEKKKEBKPTFDVSKKKNPQVNHSQLN	729	
Qy	121	ESHRKEDLQREHSQKSDSTKDVTATVLDKNKISSKSTNNPNK	164	
Db	730	ESHRKEDLQREHSQKSDSTKDVTATVLDKNKISSKSTNNPNK	773	

RESULT 2  
US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; FOR DIAG

STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
PRIORITY APPLICATION NUMBER: 60/085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 637 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATON 1...637  
SEQUENCE DESCRIPTION: SEQ ID NO: 3169:  
US-09-107-433-3169  
Query Match 99.6%; Score 845; DB 2; Length 637;  
Best Local Similarity 99.4%; Pred. No. 9.8e-79;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTAKFELNKDGTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60  
DB 440 TTAKFELNKDGTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 499  
QY 61 WEISGEGKKGAGYVNLKDTFKVPFKIEEKEEENKPTFDVSKKKNPQVNHSQLN 120  
DB 500 WEISGEGKKGAGYVNLKDTFKVPFKIEEKEEENKPTFDVSKKKNPQVNHSQLN 559  
QY 121 ESHRKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164  
DB 560 ESHRKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 603  
RESULT 3  
US-09-583-110-5274  
Sequence 5274, Application US/09583110  
Patent No. 6699703  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al.  
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics  
FILE REFERENCE: PATH00-07A  
CURRENT APPLICATION NUMBER: US/09/583,110  
CURRENT FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: US 09/107,433  
PRIORITY FILING DATE: 1998-06-30  
PRIORITY APPLICATION NUMBER: US 60/085,131  
PRIORITY FILING DATE: 1998-05-12  
PRIORITY APPLICATION NUMBER: US 60/051,553

PRIOR FILING DATE: 1997-07-02  
NUMBER OF SEQ ID NOS: 5322  
SEQ ID NO 5274  
LENGTH: 2138  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-583-110-5274  
Query Match 99.6%; Score 845; DB 2; Length 2138;  
Best Local Similarity 99.4%; Pred. No. 5.2e-78;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTAKFELNKDGTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60  
DB 1941 TTAKFELNKDGTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2000  
QY 61 WEISGEGKKGAGYVNLKDTFKVPFKIEEKEEENKPTFDVSKKKNPQVNHSQLN 120  
DB 2001 WEISGEGKKGAGYVNLKDTFKVPFKIEEKEEENKPTFDVSKKKNPQVNHSQLN 2060  
QY 121 ESHRKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164  
DB 2061 ESHRKEDLQREHHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2104  
RESULT 4  
US-08-961-083-68  
Sequence 68, Application US/08961083  
Patent No. 6159469  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESS: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,083  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340F2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-961-083-68  
Query Match 72.5%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 5.6e-56;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 YKGELEKGYQFDGWEISGEGKKGAGYVNLKDTFKVPFKIEEKEEENKPTFDVSK 107

```
Db 1 YKGELEKGYQPDGWEISGFEKGKQAGYVNLNLSKDTFKPVPFKIEKKKEENKPTFDVSK 60
Qy 108 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 164
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 117

RESULT 5
US-09-536-784-68
; Sequence 68, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match 72.5%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.6e-56;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 YKGELEKGYQPDGWEISGFEKGKQAGYVNLNLSKDTFKPVPFKIEKKKEENKPTFDVSK 107
Db 1 YKGELEKGYQPDGWEISGFEKGKQAGYVNLNLSKDTFKPVPFKIEKKKEENKPTFDVSK 60
Qy 108 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 164
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 117

RESULT 6
US-09-765-271-68
; Sequence 68, Application US/09765271
; Patent No. 6887663
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452

Db 1 YKGELEKGYQPDGWEISGFEKGKQAGYVNLNLSKDTFKPVPFKIEKKKEENKPTFDVSK 60
Qy 108 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 164
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 117

RESULT 7
US-09-765-272A-68
; Sequence 68, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
; Vaccines
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
```

```
/ APPLICATION NUMBER: US/09/765,272A
/ FILING DATE: 22-Jan-2001
/ CLASSIFICATION DATA: <Unknown>
/ PRIOR APPLICATION NUMBER: 08/961,083
/ FILING DATE: OCT-30-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Lin J. Hymel
/ REGISTRATION NUMBER: 45,414
/ REFERENCE/DOCKET NUMBER: PB340P2C2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 610-5790
/ TELEFAX: (301) 309-8439
/ INFORMATION FOR SEQ ID NO: 68:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 117 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272A-68

Query Match          72.5%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.6e-56;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YGELEKGYQPDGWEISGPEGKKGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 107
Db 1 YGELEKGYQPDGWEISGPEGKKGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 60

QY 108 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 8
US-09-710-279-652
; Sequence 652, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 652
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-652

Query Match          14.0%; Score 119; DB 2; Length 746;
Best Local Similarity 27.0%; Pred. No. 0.0013;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYK-----GELEKGYQPDGW-- 61
Db 585 EDSVNAQSLKP-----ITIGNKQIKQSQVSGTKVLPKSHKVMMLTMDGELTMP-DMTGWTK 639

QY 62 -RISGPE-----GKKDAGYVIN--LSKDTFIKPVFKIIEKKEENKPTFDVS---K 107
Db 640 EDVLAFEDLTUKVSTKGNGFVNTQSIKQIILK-----NKKIEVLSLAE 686

QY 108 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 161
Db 108 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 161
```

```
Db 687 TDDQEKTDDESDNKSCKKOKADEHDSNTSSSTKN-----DKSNADSKNDSDD 734

RESULT 9
US-09-134-001C-3868
; Sequence 3868, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,179
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3868
; LENGTH: 778
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3868

Query Match          14.0%; Score 119; DB 2; Length 778;
Best Local Similarity 27.0%; Pred. No. 0.0014;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYK-----GELEKGYQPDGW-- 61
Db 617 EDSVNAQSLKP-----ITIGNKQIKQSQVSGTKVLPKSHKVMMLTMDGELTMP-DMTGWTK 671

QY 62 -RISGPE-----GKKDAGYVIN--LSKDTFIKPVFKIIEKKEENKPTFDVS---K 107
Db 672 EDVLAFEDLTUKVSTKGNGFVNTQSIKQIILK-----NKKIEVLSLAE 718

QY 108 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 161
Db 719 TDDQEKTDDESDNKSCKKOKADEHDSNTSSSTKN-----DKSNADSKNDSDD 766

RESULT 10
US-09-248-796A-16224
; Sequence 16224, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16224
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16224

Query Match          12.0%; Score 101.5; DB 2; Length 347;
Best Local Similarity 21.4%; Pred. No. 0.029;
Matches 31; Conservative 32; Mismatches 63; Indels 19; Gaps 4;

QY 32 EMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGF-----EGKKGAGYVNLKSDTFIK 85
Db 202 KLSPLMIDEITPIEYND--KKWYVVTATSLQVYQVTDLESSRSEIGWEDDLEENYRTG 259

QY 86 PVFKKIEKKEENKPTFDVSKKCNPNVNHSQLN-----ESHRKDLQREHSQKSDS 139
```

```

; OTHER INFORMATION: Polypeptide Accession Number P46821
US-09-538-092-1135

Query Match
Best Local Similarity 11.5%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKDTGVSSELKPHRVTTVTIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI 63
Db 584 EKVMVKKDKPVKTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623

QY 64 SGFEGKKDAGYVINLSKDTPIKPVFKKIEEKKKEENKPTFDVSKKKDNPNQVNHSQLNESH 123
Db 624 QATDVKPKAAKKTVKETKVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KCEBKP 675

QY 124 RKEDLQRE-----EHSQKSDSTKDV 143
Db 676 KKEEVKKEVKKEIKKEEKKKEPKKEV 700

RESULT 13
US-09-949-016-10237
; Sequence 10237, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10237
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10237

Query Match
Best Local Similarity 11.5%; Score 97.5; DB 2; Length 2522;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKDTGVSSELKPHRVTTVTIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI 63
Db 638 EKVMVKKDKPVKTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 677

QY 64 SGFEGKKDAGYVINLSKDTPIKPVFKKIEEKKKEENKPTFDVSKKKDNPNQVNHSQLNESH 123
Db 678 QATDVKPKAAKKTVKETKVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KCEBKP 729

QY 124 RKEDLQRE-----EHSQKSDSTKDV 143
Db 730 KKEEVKKEVKKEIKKEEKKKEPKKEV 754

RESULT 14
US-09-538-092-1316
; Sequence 1316, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
```

```

; OTHER INFORMATION: Polypeptide Accession Number P46821
US-09-538-092-1135

Query Match
Best Local Similarity 11.5%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKDTGVSSELKPHRVTTVTIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI 63
Db 584 EKVMVKKDKPVKTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623

QY 64 SGFEGKKDAGYVINLSKDTPIKPVFKKIEEKKKEENKPTFDVSKKKDNPNQVNHSQLNESH 123
Db 624 QATDVKPKAAKKTVKETKVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KCEBKP 675

QY 124 RKEDLQRE-----EHSQKSDSTKDV 143
Db 676 KKEEVKKEVKKEIKKEEKKKEPKKEV 700

RESULT 12
US-09-538-092-1135
; Sequence 1135, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqformatter Version 0.9
; SEQ ID NO 1135
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
```

```

; OTHER INFORMATION: Polypeptide Accession Number P46821
US-09-538-092-1135

Query Match
Best Local Similarity 11.5%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKDTGVSSELKPHRVTTVTIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI 63
Db 584 EKVMVKKDKPVKTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623

QY 64 SGFEGKKDAGYVINLSKDTPIKPVFKKIEEKKKEENKPTFDVSKKKDNPNQVNHSQLNESH 123
Db 624 QATDVKPKAAKKTVKETKVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KCEBKP 675

QY 124 RKEDLQRE-----EHSQKSDSTKDV 143
Db 676 KKEEVKKEVKKEIKKEEKKKEPKKEV 700

RESULT 12
US-09-538-092-1135
; Sequence 1135, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqformatter Version 0.9
; SEQ ID NO 1135
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
```

; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormat Version 0.9  
; SEQ ID NO 1316  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)..(0)  
; OTHER INFORMATION: Polypeptide Accession Number Q14093  
US-09-538-092-1316

Query Match 11.2%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 0.13;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;  
QY 49 KGELEKGYQFDGWEISGFGKQDAGYVNLKDTIKPVFKIIEKKEENKPTP---DV 105  
Db 205 ESEGEKG-----GTEKDSKKGKDS-----KKGKDSAIELQAVKADKDEGKKDKANGDE 256  
QY 106 SK--KKDNPOVNHSLN-----BSHRKEDLQREHSOKSDTKD---VTATVLDKNNI 153  
Db 257 SKDAKDAKEIKKGKDKKKPSTSDSKDDVKKE---SKDRTKDAKVAKOTKEKESA 313  
QY 154 SSK 156  
Db 314 DSK 316

RESULT 15  
US-10-172-502-10  
; Sequence 10, Application US/10172502  
; Patent No. 6841154  
; GENERAL INFORMATION:  
; APPLICANT: FOSTER, Timothy et al.  
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .  
; FILE REFERENCE: P07263US01/BAS  
; CURRENT APPLICATION NUMBER: US/10/172,502  
; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: US 60/298,098  
; PRIOR FILING DATE: 2001-06-15  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 654  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-172-502-10

Query Match 11.0%; Score 93.5; DB 2; Length 654;  
Best Local Similarity 21.9%; Pred. No. 0.46;  
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;  
QY 1 TTVKEFILNKDTGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 56  
Db 336 SAITEFQNVQPTNEKMTLDQTKVYVESVNNESMMDTFVKH-----PIKGTMLNGKKY 390  
QY 57 QP-----DGWEISGFGKK-----D 71  
Db 391 MYMETNDNDYDKDFMVEGQVRTISKAKNRTTIFPVYEGKTLVDALVKVHVKTIDYD 450  
QY 72 AGVINL-SKDTFIKPVFKIIEKKEENKPTFDV-----SKKKNPQVNHSLQNESHRK 125  
Db 451 GQYHVIVDKAFTKANTDKSNKGEQDNSAKKEATPATPSTPSVPEKESQKQDSQKD 510  
QY 126 EDLQ-----REHSQKSDSTKQVT--ATVLDKNNISSKSTTNNPNK 164  
Db 511 DNKQLPSVKEKNDASSGSKDKTPATKPTKGEVSSSTT--PTK 552

## RESULT 16

US-09-248-796A-17646  
; Sequence 17646, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 17646  
; LENGTH: 280  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-17646

Query Match 11.0%; Score 93; DB 2; Length 280;  
Best Local Similarity 25.7%; Pred. No. 0.16;  
Matches 43; Conservative 26; Mismatches 58; Indels 40; Gaps 10;  
QY 8 LNKDTGEVSELKPHRVTVTIQNGKEMMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFE 67  
Db 67 INPIVGHVSS-----TYTVKT--SVASTFCSDYDFNVFSYASNLGLF-----ELSYA 113  
QY 68 GKDDAGYVNLKDTFIKPVFKIE-EKKEENKPTFDVSKKKNPQVNHSLQNESHRK 125  
Db 114 NKK-----KNSF--PSFEHHEIHSSSEENK-----YLKQHPQLQRRHNLHNLHQR 158  
QY 126 EDLQREHS-----QKSDSTKQVTATVLDKNNISSKSTTNNPN 163  
Db 159 VPIKSHKYEGNRTIPIQLNDNVYINPTLLSSNG-STSTTTNNEN 204

## RESULT 17

US-09-710-279-1888  
; Sequence 1888, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P03480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1888  
; LENGTH: 299  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-1888

Query Match 10.8%; Score 92; DB 2; Length 299;  
Best Local Similarity 24.4%; Pred. No. 0.22;  
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;  
QY 9 NKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFI----LPVYKGELEKG-YQFDGWEI 63  
Db 33 NKDT-EKSDKKYHRIISLIPSTHEILYRLGIGEDIVGVSTVDDYDKVKKKKQFDANL 91  
QY 64 SGFE-----GKKDAGYVNLKDTFIKPV-----FKKIEKK 95

Db 92 NKEELIKAKPDLILAHESQKNSAGKVLKSLKDKGVKVVVKDAQSIDETDTFKSIGQLT 151  
QY 96 ERENKPTFDVSKKDNQPNVHSQLNESHKEDLQREHSQ 135  
Db 152 DREKQAKELVDTEKENVEKIINSVPKHKKQEVFMEVSSK 191

## RESULT 18

US-09-134-001C-5667  
; Sequence 5667, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5667  
; LENGTH: 309  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5667

Query Match 10.8%; Score 92; DB 2; Length 309;  
Best Local Similarity 24.4%; Pred. No. 0.23; Mismatches 60; Indels 34; Gaps 5;  
Matches 39; Conservative 27;  
QY 9 NKDTGEVSELKPHRVTTTQNGKEMSSITVSEEDFI----LPVYKGELEKG-YQPDGWEI 63  
Db 43 NKDT-EKSDKKVRIISLIPSNTEILYRLGIGEDIVGVSTVDYDFKDVKKGKQFDANWL 101  
QY 64 SGFE-----GKQDAGYVNLKSDTFIKPV-----FKKIBKK 95  
Db 102 NKEELIKAKPDLILAHESQKNSAGKVLKSLKDKGVKVVVKDAQSIDETDTFKSIGQLT 161  
QY 96 ERENKPTFDVSKKDNQPNVHSQLNESHKEDLQREHSQ 135  
Db 162 DREKQAKELVDTEKENVEKIINSVPKHKKQEVFMEVSSK 201

## RESULT 19

US-09-134-001C-3033  
; Sequence 3033, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3033  
; LENGTH: 442  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3033

Query Match 10.7%; Score 91; DB 2; Length 442;  
Best Local Similarity 21.7%; Pred. No. 0.49;  
Matches 49; Conservative 30; Mismatches 65; Indels 82; Gaps 11;  
QY 10 KDTGEVSELKPHRVTTT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 54

Db 90 EDT-EINVAQPEVNVNTQIEKGKDFIFEATVTVPEVKLGDKYKGLEIEKQETDLSDEBLQ 148  
QY 55 -----GYQPDGWEISG--PEGKKDAGVWVNLKSDTFIK 85  
Db 149 ESIDHSLSLAEMVVKEDGAVENGDTVNIIDFSG-SVDGEEDFGGQAGYDLEIGSGSIFP 207  
QY 86 PVFKKIBKKIBKBNK-----PTFDVS-----KKDNQPNVHSQLNE- 121  
Db 208 GFEEQIEGMKTGDEKDVVVVTFPEYHABELAGKEATFKTVNEIKFKDVPDLNDEIANEL 267  
QY 122 -----SHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNN 160  
Db 268 DSDAENVDEYKENTLRKLSLSEQKATEAENT-----EKERAINKATEN 308

## RESULT 20

US-09-134-001C-5157  
; Sequence 5157, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5157  
; LENGTH: 902  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5157

Query Match 10.7%; Score 91; DB 2; Length 902;  
Best Local Similarity 24.9%; Pred. No. 1.3; Mismatches 31; Indels 24; Gaps 8;  
Matches 43; Conservative 31;  
QY 1 TTVKEFILKDTGVESELKPHRVTTTQNGKEMS-STIVSEEDFILPVYKGELEKGYQFD 59  
Db 736 TPVSEYRLSNRGGKIK----TATITERNGNVICITTVTGEEDLWVVTNAGVI---IRLD 788  
QY 60 GWEISGPEGKKDAGY-VNLKSDTFIKPVFKKIBKKEEN-----KPTFDVSKK 109  
Db 789 VHDISQ-NGRAAQGVRLMKLGDGQFVSTVAKVNEEDDNEENADEAQOSTTTTADVEEVV 847  
QY 110 DNPQVHSQLNESHKEDLQREH-HSQKSDSTKDVATVLDKNNISSKSTNN 161  
Db 848 D----DQTGPAIHTEGDAEMESVESPENDDRIDIRQDPMDRVNEDIESASN 896

## RESULT 21

US-08-910-925-3  
; Sequence 3, Application US/08910925  
; Patent No. 6162601  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:





```
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 21451
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-21451

Query Match
Best Local Similarity 10.6%; Score 89.5; DB 2; Length 262;
Matches 43; Conservative 26; Mismatches 43; Indels 53; Gaps 9;

QY 10 KDTGEVSEL-KHRRVTVT-----IQNGKMSSTIVSEEDP-----43
Db 52 KSTPKTSPURKPKPKPTVTVRKWASKRPPSVTNTPEIKPKSSSPFIISSEDFBLEWMD 111
QY 44 ----ILPVYKGELEKGYQDFGWEISGFEGKDGAGVYVNLKSDTKFVKPKIEKKKEEN 99
Db 112 KLTEEVPIIE-----FNDNYD--NDEKKKE--VVKSKKE-NQNTKGHELKPKPKP 160
QY 100 KPTFDVSKKKDQPNVNHSQLNSHREKEDLQREHSQKSDSTKDV 144
Db 161 KPSEETSKTKVKPQ-----PRK---QKKQKPLSEETVDTLT 193

RESULT 25
US-09-464-483-4
; Sequence 4, Application US/09464483
; Patent No. 6228617
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6228617el tig
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/464,483
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,339
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-464-483-4

Query Match
Best Local Similarity 10.6%; Score 89.5; DB 2; Length 402;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 11 DTGEVSELKPHRRVTVT-IQNGKE--MSSTIVSEEDFILPVYKGBLEK-----54
Db 53 DETDIKPVAQPEVSVTQIEKGKDFIFEATVTVPEPVKGLGYKGLSIEKQETLSDELOE 112
QY 55 -----GYQDFGWEISG--FEGKDGAGVYVNLKSDTKFVKP 86
Db 113 AIDHSLGHLAEMVVKEDGVVNGDVTNIDFSG--VDGSEFEGGQAGYDLEIGSGSFI-P 170
QY 87 VPK-----KIEKKKE-----EE--NKPTFDVS-----KKONPQNVHSQLNE- 121
Db 171 GFEEQLEGKMKVDEEKDVVVVTPPEYHABELAGKEATFKTKVNEIKFKEVPELTDEIANEL 230

US-09-414-664-4
; Sequence 4, Application US/09414664
; Patent No. 6242249
; GENERAL INFORMATION:
; APPLICANT: Lawlor, Elizabeth J.
; TITLE OF INVENTION: No. 6242249el tig
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/414,664
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/999,339
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-414-664-4

Query Match
Best Local Similarity 10.6%; Score 89.5; DB 2; Length 402;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 11 DTGEVSELKPHRRVTVT-IQNGKE--MSSTIVSEEDFILPVYKGBLEK-----54
Db 53 DETDIKPVAQPEVSVTQIEKGKDFIFEATVTVPEPVKGLGYKGLSIEKQETLSDELOE 112
QY 55 -----GYQDFGWEISG--FEGKDGAGVYVNLKSDTKFVKP 86
Db 113 AIDHSLGHLAEMVVKEDGVVNGDVTNIDFSG--VDGSEFEGGQAGYDLEIGSGSFI-P 170
QY 87 VPK-----KIEKKKE-----EE--NKPTFDVS-----KKONPQNVHSQLNE- 121
Db 171 GFEEQLEGKMKVDEEKDVVVVTPPEYHABELAGKEATFKTKVNEIKFKEVPELTDEIANEL 230
```





```
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Bidhinn, Deirdre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1315
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-200-650E-5

Query Match          10.3%; Score 87.5; DB 2; Length 1315;
Best Local Similarity 25.0%; Pred. No. 5;
Matches 41; Conservative 23; Mismatches 73; Indels 27; Gaps 5;

QY 9 NKDTGEVSELPK-HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQDFGWSISGF- 66
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 804 NKDGKQDSTKGIQGVTVTLKN-----ENGEVLQTTKTDKDGKYQFTGLENGTYK 853
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

QY 67 -EGKXGAGVNLSDTKFKPVFKIEKKEENKPTFDVSKKDNQVNHSQLNESHK 125
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 854 VEFETPSGYTPT-----QVSGTGDEGDSNGSTSTTGVIKDKNDRTIDSGFYKPTNKL 905
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

QY 126 EDLQREHSHQSDSTKD-----VTATVLDKNNISSKSTNNPN 163
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 906 GDVWVEDTKNGVDKDEKIGSVTVTLKENDKVLKVTITDEN 949
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

RESULT 34
US-08-296-791-5
; Sequence 5, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard P.
; REGISTRATION NUMBER: 31,801
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard P.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
```

```
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-08-296-791-5

Query Match          10.3%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 7.1;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 76 INLSKDTFKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSH- 134
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1296 INTGSATAITETAESKDPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

QY 135 --OKSDSTKQVTVATVLDKNNISSKSTNNPNK 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1356 SQPQTSABETTAASTDETTIADNSKSKPNR 1387
   : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 35
US-09-839-996-5
; Sequence 5, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard P.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
; US-09-839-996-5

Query Match          10.3%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 7.1;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 76 INLSKDTFKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSH- 134
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 1296 INTGSATAITETAESKDPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
```

QY	135	--QKSDTQVATVLDKNNISSKSTNNPNK 164
Db	1356	SQOQTSABETTAASDTETTTIADNSKRSKPNR 1387
RESULT 36		
US-10-080-505-5		
Sequence 5, Application US/10080505		
Patent No. 6676948		
GENERAL INFORMATION:		
APPLICANT: St. Geme, Joseph W.		
TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS		
FILE REFERENCE: A-59941-1/RPT/DCF/DHR		
CURRENT APPLICATION NUMBER: US/10/080,505		
CURRENT FILING DATE: 2002-02-22		
PRIOR APPLICATION NUMBER: US 08/296,791		
PRIOR FILING DATE: 1994-10-25		
PRIOR APPLICATION NUMBER: US 09/839,996		
PRIOR FILING DATE: 2001-04-20		
NUMBER OF SEQ ID NOS: 58		
SOFTWARE: Patent in version 3.1		
SEQ ID NO 5		
LENGTH: 1702		
TYPE: PRT		
ORGANISM: Haemophilus influenzae		
US-10-080-505-5		
Query Match	10.3%;	Score 87.5; DB 2; Length 1702;
Best Local Similarity	27.2%;	Pred. No. 7.1;
Matches	25; Conservative	14; Mismatches 50; Indels 3; Gaps 1;
QY	76	INLSKDTFKPVFKKIEBKKEENKPTFDVSKKONQVNHSQLNESHKEDLQREHS- 134
Db	1296	INTGSATAITETAESKDPQTETAASDASQHKANTVADNSVANNSESSEPKSRRSI 1355
QY	135	--QKSDTQVATVLDKNNISSKSTNNPNK 164
Db	1356	SQOQTSABETTAASDTETTTIADNSKRSKPNR 1387
RESULT 37		
US-10-080-505-5		
Sequence 5, Application US/10645655		
Patent No. 6815182		
GENERAL INFORMATION:		
APPLICANT: St. Geme III, Joseph W.		
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein		
NUMBER OF SEQUENCES: 9		
CORRESPONDENCE ADDRESS:		
ADDRESSER: Flehr, Hobbach, Test, Albritton & Herbert		
STREET: 4 Embarcadero Center, Suite 3400		
CITY: San Francisco		
STATE: California		
COUNTRY: United States		
ZIP: 94111-4187		
COMPUTER READABLE FORM:		
MEDIUM TYPE: Floppy disk		
COMPUTER: IBM PC compatible		
OPERATING SYSTEM: PC-DOS/MS-DOS		
SOFTWARE: Patent in Release #1.0, Version #1.25		
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: PCT/US95/10661A		
FILING DATE: 16-AUG-1995		
CLASSIFICATION:		
PRIOR APPLICATION DATA:		
APPLICATION NUMBER: US 08/296,791		
FILING DATE: 25-AUG-1994		
CLASSIFICATION:		
ATTORNEY/AGENT INFORMATION:		
NAME: Trecartin, Richard F.		
REGISTRATION NUMBER: 31,801		
REFERENCE/DOCKET NUMBER: FP-59941/RFT		
TELECOMMUNICATION INFORMATION:		
TELEPHONE: (415) 781-1989		
TELEFAX: (415) 398-3249		
TELEX: 910 277299		
INFORMATION FOR SEQ ID NO: 5:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 1702 amino acids		
TYPE: amino acid		
TOPOLOGY: unknown		
PCT-US95-10661A-5		
Query Match	10.3%;	Score 87.5; DB 4; Length 1702;
Best Local Similarity	27.2%;	Pred. No. 7.1;
Matches	25; Conservative	14; Mismatches 50; Indels 3; Gaps 1;
QY	76	INLSKDTFKPVFKKIEBKKEENKPTFDVSKKONQVNHSQLNESHKEDLQREHS- 134
Db	1296	INTGSATAITETAESKDPQTETAASDASQHKANTVADNSVANNSESSEPKSRRSI 1355
QY	135	--QKSDTQVATVLDKNNISSKSTNNPNK 164
Db	1356	SQOQTSABETTAASDTETTTIADNSKRSKPNR 1387
RESULT 38		
PCT-US95-10661A-5		
Sequence 5, Application PC/TUS9510661A		
GENERAL INFORMATION:		
APPLICANT: Washington University, et al.		
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein		
NUMBER OF SEQUENCES: 9		
CORRESPONDENCE ADDRESS:		
ADDRESSER: Flehr, Hobbach, Test, Albritton & Herbert		
STREET: 4 Embarcadero Center, Suite 3400		
CITY: San Francisco		
STATE: California		
COUNTRY: United States		
ZIP: 94111-4187		
COMPUTER READABLE FORM:		
MEDIUM TYPE: Floppy disk		
COMPUTER: IBM PC compatible		
OPERATING SYSTEM: PC-DOS/MS-DOS		
SOFTWARE: Patent in Release #1.0, Version #1.25		
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: PCT/US95/10661A		
FILING DATE: 16-AUG-1995		
CLASSIFICATION:		
PRIOR APPLICATION DATA:		
APPLICATION NUMBER: US 08/296,791		
FILING DATE: 25-AUG-1994		
CLASSIFICATION:		
ATTORNEY/AGENT INFORMATION:		
NAME: Trecartin, Richard F.		
REGISTRATION NUMBER: 31,801		
REFERENCE/DOCKET NUMBER: FP-59941/RFT		
TELECOMMUNICATION INFORMATION:		
TELEPHONE: (415) 781-1989		
TELEFAX: (415) 398-3249		
TELEX: 910 277299		
INFORMATION FOR SEQ ID NO: 5:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 1702 amino acids		
TYPE: amino acid		
TOPOLOGY: unknown		
PCT-US95-10661A-5		
Query Match	10.3%;	Score 87.5; DB 4; Length 1702;
Best Local Similarity	27.2%;	Pred. No. 7.1;
Matches	25; Conservative	14; Mismatches 50; Indels 3; Gaps 1;
QY	76	INLSKDTFKPVFKKIEBKKEENKPTFDVSKKONQVNHSQLNESHKEDLQREHS- 134

Db 1296 INTGSAITETAEKSDRQDTAATASQHKANTVADNSVANSSESPKSRRSI 1355  
QY 135 --QKSDSTKDVATVLDKNNISSKSTNNPNK 164  
Db 1356 SQOQTSABETTAASTDETTIADNSKRSPNR 1387

RESULT 39

US-09-830-230A-651  
; Sequence 651, Application US/09830230A  
; Patent No. 6902893  
; GENERAL INFORMATION:  
; APPLICANT: Human Genome Sciences, Inc.  
; TITLE OF INVENTION: Lyme Disease Vaccines  
; FILE REFERENCE: PB481US  
; CURRENT APPLICATION NUMBER: US/09/830,230A  
; CURRENT FILING DATE: 2001-09-27  
; PRIOR APPLICATION NUMBER: PCT/US98/12718  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/057,483  
; PRIOR FILING DATE: 1997-09-03  
; PRIOR APPLICATION NUMBER: 60/053,344  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/053,377  
; PRIOR FILING DATE: 1997-07-22  
; PRIOR APPLICATION NUMBER: 60/050,359  
; PRIOR FILING DATE: 1997-06-20  
; NUMBER OF SEQ ID NOS: 756  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 651  
; LENGTH: 277  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-830-230A-651

Query Match 10.2%; Score 86.5; DB 2; Length 277;  
Best Local Similarity 20.8%; Pred. No. 0.74; Indels 57; Gaps 4;  
Matches 32; Conservative 26; Mismatches 57; Indels 39; Gaps 4;  
QY 3 VKFEILNKDTGVSSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWE 62  
Db 151 IENFFQNDLLFVLTKDKNNNTINIMLPNDIQPKDYILKDKTIKKG----- 203  
QY 63 ISGFEGKQAGVYINLSKDTFTKPVFKIEEKKBEENKPTFDVSKKKDNPNQVNHSQLNES 122  
Db 204 ----TGEK-----YLNPIYR-----FQIKNKDYHSIDYNKVTIS 234  
QY 123 HRKEDLQREHSQKSDSTKDVTA---TVLDKNNI 153  
Db 235 EKTIELDLPHPQVFMKNFTKILDTITDLNNL 268

RESULT 40

US-08-533-669A-18  
; Sequence 18, Application US/08533669A  
; Patent No. 5834592  
; GENERAL INFORMATION:  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/533,669A  
; FILING DATE: 22-SEP-1995  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 732 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-533-669A-18  
Query Match 10.2%; Score 86.5; DB 1; Length 732;  
Best Local Similarity 23.6%; Pred. No. 2.8; Indels 39; Gaps 8;  
Matches 39; Conservative 29; Mismatches 58; Indels 39; Gaps 8;  
QY 6 FILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65  
Db 170 FTVRTDTGE-----PMGRGTVLHLKEDQTEYLSERRI-----KEIVKKSQFIGYPITL 220  
QY 66 F-EKKQDAGVYINLSKDTFTKPVFKIEEKKBEENKPTFDVSKKKDNPNQVNHSQLNESHR 124  
Db 221 FVEKERDK-----EVSDD-----EAEKEKDEEKEKEKESEDKPEI----- 258  
QY 125 KEDLQREHSQKSD-----STKDVTAATVLDKNNISKST--TNNPN 163  
Db 259 -EDVGSDEEDEKDGDKKKKKKKIKKIKYIDKEELNKTPIWTRNPD 302  
Search completed: April 24, 2006, 15:03:39  
Job time : 26.4989 secs

Result No.	Score	Query			DB	ID	Description
		Match	Length	†			
1	897	100.0	773	4	AA848343	Aa848343 S. pneumo	
2	897	100.0	2120	3	AAy81710	Aay81710 Streptoco	
3	897	100.0	2140	6	ABU01020	Abu01020 S. pneumo	
4	897	100.0	2140	6	ADu5746	Adu5746 Protein e	
5	897	100.0	2140	8	ADW92113	Adw92113 S. pneumon	
6	897	100.0	2140	8	ADT50039	Adt50039 S. pneumon	
7	894	99.7	637	8	ADR94534	Adr94534 Novel S.	
8	894	99.7	637	9	AEa58404	Aea58404 Streptoco	
9	894	99.7	2138	8	ADK48759	Adk48759 Streptoco	
10	615	68.6	117	2	AAW55036	Aaw55036 Streptoco	
11	615	68.6	117	5	ABp54590	Abp54590 S. pneumo	
12	615	68.6	117	7	ADC45149	Adc45149 S. pneumo	
13	121	13.5	778	5	ABp39023	Abp39023 Staphyloc	
14	121	13.5	778	8	ADSp6368	Adsp6368 Staphyloc	
15	119	13.3	746	4	AAg81779	Aag81779 S. epider	
16	112.5	12.5	354	9	ADz72253	Adz72253 Plasmodiu	
17	111.5	12.4	707	6	ABu25018	Abu25018 Protein e	
18	111.5	12.4	775	6	ABu42797	Abu42797 Protein e	
19	110.5	12.3	647	9	ADZ79635	Adz79635 P. falcip	
20	110.5	12.3	651	8	ADo19012	Ado19012 Amino aci	
21	110.5	12.3	651	8	ADo119010	Ado119010 P. falcip	
22	109	12.2	665	3	AAb18278	Aab18278 Plasmodiu	
23	109	12.2	665	7	ABo23606	AbO23606 Plasmodiu	
24	108	12.0	188	9	ADZ79639	Adz79639 P. falcip	

CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Sp130 polypeptide  
XX  
SQ Sequence 773 AA;

Query Match 100.0%; Score 897; DB 4; Length 773;  
Best Local Similarity 100.0%; Pred. No. 6.6e-77;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKEFILNKDTGVESELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 60  
DB 600 KIVVKDFARNTTVKEFILNKDTGVESELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 659  
QY 61 ELEKGYPDQGWELSGPEGKDGAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKD 120  
DB 660 ELEKGYPDQGWELSGPEGKDGAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKD 719  
QY 121 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 174  
DB 720 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 773

RESULT 2  
AY81710  
ID AAY81710 standard; protein; 2120 AA.  
XX  
AC AAY81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PP 27-JUL-1999; 99WO-GB002452.  
XX  
PR 27-JUL-1998; 98GB-00016336.  
PR 19-MAR-1999; 99US-0125329P.  
XX  
PA (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
PI Le Page RWP, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
DR WPI; 2000-195301/17.  
DR N-PSDB; AAZ91806.  
XX  
PT Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections.  
XX  
PS Claim 2; Page 41-42; 76pp; English.  
XX  
CC This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis  
XX  
SQ Sequence 2120 AA;

Query Match 100.0%; Score 897; DB 3; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 2.6e-76;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKEFILNKDTGVESELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 60  
DB 1913 KIVVKDFARNTTVKEFILNKDTGVESELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 1972  
QY 61 ELEKGYPDQGWELSGPEGKDGAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKD 120  
DB 1973 ELEKGYPDQGWELSGPEGKDGAGVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKD 2032  
QY 121 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 174  
DB 2033 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPNK 2086

RESULT 3  
ABU01020  
ID ABU01020 standard; protein; 2140 AA.  
XX  
AC ABU01020;  
XX  
DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX  
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KW gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
PN WO200277021-A2.  
XX  
PD 03-OCT-2002.  
XX  
PP 27-MAR-2002; 2002WO-IB002163.  
XX  
PR 27-MAR-2001; 2001GB-00007658.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Masignani V, Tettelin H, Fraser C;  
XX  
DR WPI; 2003-040579/03.  
DR N-PSDB; ABX06302.  
XX  
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
PS Claim 1; SEQ ID NO 1180; 56pp; English.  
XX  
CC The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC ABS56454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target



CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 897; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-76;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKG 60  
 DB 1933 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKG 1992

QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTPIKPVFKKIBKKEENKPTFDVSKKKD 120  
 DB 1993 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTPIKPVFKKIBKKEENKPTFDVSKKKD 2052

QY 121 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 174  
 DB 2053 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 2106

RESULT 4  
 ABU45746  
 ID ABU45746 standard; protein; 2140 AA.  
 AC ABU45746;  
 XX  
 XX  
 XX  
 XX 19-JUN-2003 (first entry)  
 XX Protein encoded by Prokaryotic essential gene #31273.  
 DE  
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 KW  
 XX Streptococcus pneumoniae.  
 OS  
 XX WO200277183-A2.  
 PN  
 XX 03-OCT-2002.  
 XX  
 XX 21-MAR-2002; 2002WO-US009107.  
 PF  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 PA  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 DR N-PSDB; ACA49616.  
 DR

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 73670; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway;  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 897; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-76;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKG 60  
 DB 1933 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKG 1992

QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTPIKPVFKKIBKKEENKPTFDVSKKKD 120  
 DB 1993 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTPIKPVFKKIBKKEENKPTFDVSKKKD 2052

QY 121 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 174  
 DB 2053 NPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 2106

RESULT 5  
 ADM92113  
 ID ADM92113 standard; protein; 2140 AA.  
 AC ADM92113;  
 XX  
 XX 03-JUN-2004 (first entry)  
 DT  
 XX S pneumoniae antigenic protein sequence SeqID310.  
 DE  
 XX antibacterial; gene therapy; Streptococcus pneumoniae infection;  
 KW antigenic.  
 XX  
 XX Streptococcus pneumoniae.  
 XX

PN WO2004020609-A2.  
XX  
XX PD 11-MAR-2004.  
XX PF 02-SEP-2003; 2003WO-US027401.  
XX PR 30-AUG-2002; 2002US-0407082P.  
XX PA (TUFT ) UNIV TUFTS.  
XX PI Camilli A, Hava DL;  
XX DR WPI; 2004-239189/22.  
XX DR N-PSDB; ADM91876.  
XX PT New Streptococcus pneumoniae nucleic acid molecules, useful for  
XX PT diagnosing, treating and preventing active infections of Streptococcus  
XX PT pneumoniae.  
XX PS Claim 27; SEQ ID NO 310; 123pp; English.  
XX CC This invention relates to novel isolated Streptococcus pneumoniae nucleic  
XX CC acid molecules and the antigenic polypeptides encoded by them. The  
XX CC invention may be useful for the production of compounds with an  
XX CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
XX CC compositions and methods disclosed are useful for treating Streptococcus  
XX CC pneumoniae infection. The present sequence is that of an S pneumoniae  
XX CC protein of the invention.  
XX CC  
XX SQ Sequence 2140 AA;  
  
Query Match 100.0%; Score 897; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.7e-76; Indels 0; Gaps 0;  
Matches 174; Conservative 0; Mismatches 0;  
  
QY 1 KIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKG 60  
DB 1933 KIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKG 1992  
  
QY 61 ELEKGYPDQGWISGFEKGKQAGYVNLKDTFKPVFKKIEKKKEENKPTFDVSKKKD 120  
DB 1993 ELEKGYPDQGWISGFEKGKQAGYVNLKDTFKPVFKKIEKKKEENKPTFDVSKKKD 2052  
  
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 174  
DB 2053 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 2106  
  
RESULT 6  
ID ADT50099  
XX ADT50099 standard; protein; 2140 AA.  
XX AC ADT50099;  
XX DT 13-JAN-2005 (first entry)  
XX DE S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
XX KW hyperimmune serum reactive antigen; antibacterial; vaccine;  
XX KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
XX KW sepsis; meningitis.  
XX OS Streptococcus pneumoniae TIGR4.  
XX PN WO2004092209-A2.  
XX PD 28-OCT-2004.  
XX PF 15-APR-2004; 2004WO-EP003984.  
XX PR 15-APR-2003; 2003EP-00450087.  
XX PA (INTE-) INTERCELL AG.

XX PI Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;  
XX DR WPI; 2004-758335/74.  
XX DR N-PSDB; ADT49955.  
XX PT New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
XX PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
XX PT treating S. pneumoniae infections.  
XX PS Disclosure; SEQ ID NO 177; 191pp; English.  
XX CC This invention relates to novel nucleic acids encoding hyperimmune serum  
XX CC reactive antigens, or fragments derived thereof. Specifically, it refers  
XX CC to antigens selected from peptides and serum reactive epitopes that can  
XX CC be used in pharmaceutical compositions that exhibit antibacterial  
XX CC activity. The present invention describes a composition (including the  
XX CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
XX CC that is useful for manufacturing a medicament such as a vaccine, which  
XX CC can be used to treat or prevent bacterial infections, particularly S.  
XX CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
XX CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
XX CC be used for isolating, purifying and/ or identifying an interaction  
XX CC partner of the hyperimmune serum reactive antigen, as well as for  
XX CC manufacturing a functional nucleic acid selected from aptamers and  
XX CC spiegelmers or for manufacturing a functional ribonucleic acid selected  
XX CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
XX CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
XX CC of the invention.  
XX CC  
XX SQ Sequence 2140 AA;  
  
Query Match 100.0%; Score 897; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.7e-76; Indels 0; Gaps 0;  
Matches 174; Conservative 0; Mismatches 0;  
  
QY 1 KIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKG 60  
DB 1933 KIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKG 1992  
  
QY 61 ELEKGYPDQGWISGFEKGKQAGYVNLKDTFKPVFKKIEKKKEENKPTFDVSKKKD 120  
DB 1993 ELEKGYPDQGWISGFEKGKQAGYVNLKDTFKPVFKKIEKKKEENKPTFDVSKKKD 2052  
  
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 174  
DB 2053 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 2106  
  
RESULT 7  
ID ADR94534  
XX ADR94534 standard; protein; 637 AA.  
XX AC ADR94534;  
XX DT 16-DEC-2004 (first entry)  
XX DE Novel S. pneumoniae protein sequence, SEQ ID 3169.  
XX KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
XX KW bacterial infection.  
XX OS Streptococcus pneumoniae.  
XX PN US6800744-B1.  
XX PD 05-OCT-2004.  
XX PF 30-JUN-1998; 98US-00107433.  
XX PR 02-JUL-1997; 97US-0051553P.  
XX PR 12-MAY-1998; 98US-0085131P.  
XX

PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2004-697205/68.  
DR N-PSDB; ADR91931.  
DR  
XX New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 3169; 151pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366 polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridizable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae protein sequences. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.  
XX  
XX Sequence 637 AA;  
SQ  
Query Match 99.7%; Score 894; DB 8; Length 637;  
Best Local Similarity 99.4%; Pred. No. 9.8e-77;  
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 60  
Db 430 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 489  
QY 61 ELEKGYQPDGWEISGFEKGDAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 120  
Db 490 ELEKGYQPDGWEISGFEKGDAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 549  
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 174  
Db 550 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 603  
RESULT 8  
AEA58404  
ID AEA58404 standard; protein; 637 AA.  
XX  
AC AEA58404;  
XX  
XX 25-AUG-2005 (first entry)  
XX  
XX Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:3169.  
DE  
XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;  
KW vaccine.  
XX  
XX Streptococcus pneumoniae.  
OS  
XX  
XX US2005136404-A1.  
PN  
XX

PD 23-JUN-2005.  
XX  
XX 10-JUL-2003; 2003US-00617320.  
XX  
XX 02-JUL-1997; 97US-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
PR 30-JUN-1998; 98US-00107433.  
XX  
XX (DOUC/) DOUCETTE-STAMM L A.  
PA (BUSH/) BUSH D.  
XX  
XX Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2005-477576/48.  
DR N-PSDB; AEA55801.  
XX  
XX New isolated nucleic acid molecules and encoded polypeptides useful for  
PT diagnosing, preventing or treating bacterial infections, particularly  
PT Streptococcus pneumoniae infection.  
XX  
XX Claim 5; SEQ ID NO 3169; 144pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule for detecting,  
CC preventing or treating pathological conditions resulting from bacterial  
CC infection. The isolated nucleic acid comprises: (a) any of the 2603  
CC nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence  
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the  
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide  
CC sequence of at least 8 nucleotides in length, where the sequence is  
CC hybridizable to a nucleic acid having any of the nucleotide sequences in  
CC (a). Also described: (1) a recombinant expression vector comprising the  
CC above nucleic acid operably linked to a transcription regulatory element;  
CC (2) a cell comprising the recombinant expression vector; (3) producing an  
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence  
CC consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)  
CC treating a subject for S. pneumoniae infection; (6) a recombinant or  
CC substantially pure preparation of an S. pneumoniae polypeptide or its  
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;  
CC (7) a vaccine composition for preventing or treating an S. pneumoniae  
CC infection, comprising an amount of the above nucleic acid or polypeptide;  
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;  
CC (9) a computer readable medium having recorded the nucleotide sequences  
CC of AEA55236 to AEA57838; (10) a computer based system for identifying  
CC fragments of the Streptococcus genome of commercial importance. The  
CC composition and methods are useful for diagnosing, preventing or treating  
CC bacterial infections, particularly S. pneumoniae infection. The present  
CC sequence represents a S. pneumoniae ORF amino acid sequence from the  
CC present invention. Note - the sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from the USPTO web site.  
XX  
XX Sequence 637 AA;  
SQ  
Query Match 99.7%; Score 894; DB 9; Length 637;  
Best Local Similarity 99.4%; Pred. No. 9.8e-77;  
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 60  
Db 430 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 489  
QY 61 ELEKGYQPDGWEISGFEKGDAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 120  
Db 490 ELEKGYQPDGWEISGFEKGDAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKD 549  
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 174  
Db 550 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 603  
RESULT 9  
ADK48759  
ID ADK48759 standard; protein; 2138 AA.

```
XX ADK48759;
AC
XX
XX
DT 20-MAY-2004 (first entry)
DE Streptococcus pneumoniae protein, Seq ID No 5274.
XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS Streptococcus pneumoniae.
XX
XX US6699703-B1.
XX
XX 02-MAR-2004.
XX
XX 26-MAY-2000; 2000US-00583110.
XX
XX 02-JUL-1997; 97US-0051553P.
XX 12-MAY-1998; 98US-0085131P.
XX 30-JUN-1998; 98US-00107433.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX
XX WPI; 2004-212399/20.
XX N-PSDB; ADK46098.
XX
XX New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX
XX Disclosure; SEQ ID NO 5274; 301pp; English.
XX
XX The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC data for this patent did not appear in the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
XX Sequence 2138 AA;
SQ
Query Match 99.7%; Score 894; DB 8; Length 2138;
Best Local Similarity 99.4%; Pred. No. 5.2e-76;
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KIVVKDFARNTTVKEPILNKDTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKG 60
DB 1931 KIVVKDFARNTTVKEPILNKDTGEVSELKPHRVTVVTIQNGKMSSTIVSEEDFILPVYKG 1990
QY 61 ELEKGQPDGWEISGEGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSKKD 120
DB 1991 ELEKGQPDGWEISGEGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSKKD 2050
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 174
DB 2051 NPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2104
RESULT 10
AAW55096
ID AAW55096 standard; protein; 117 AA.
XX
XX AAW55096;
XX
XX 02-OCT-1998 (first entry)
XX
XX Streptococcus pneumoniae SP0043 protein.
DE
```

```
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
XX WO9818930-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US019422.
XX
XX 31-OCT-1996; 96US-0029960P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX
XX WPI; 1998-272224/24.
XX N-PSDB; AAV27357.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.
XX
XX Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose
XX Sequence 117 AA;
SQ
Query Match 68.6%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.9e-51;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 58 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 117
DB 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEENKPTFDVSK 60
QY 118 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 174
DB 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117
RESULT 11
ABP54590
ID ABP54590 standard; protein; 117 AA.
XX
XX ABP54590;
XX
XX 04-SEP-2002 (first entry)
XX
XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
XX
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
XX
XX US2002061545-A1.
PN
```

```
XX 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-00765272.
XX
XX 30-OCT-1997; 97US-00961083.
XX
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.
XX (DOUG/) DOUGHERTY B.
XX (FANN/) FANNON M R.
XX (ROSE/) ROSEN C A.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX
XX WPI; 2002-479261/51.
XX N-PSDB; ABQ84825.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX and for preventing or attenuating disease caused by Streptococcus
XX infection.
XX
XX Claim 11; Page 29; 70pp; English.
XX
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX pneumoniae antigens have antibacterial activity and can be used in
XX vaccines. The S. pneumoniae antigens can also be used to prevent or
XX attenuate a Streptococcal infection in an animal. The polynucleotides
XX encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX of S. pneumoniae ORFs (open reading frames) which are used in an example
XX from the present invention
XX
XX Sequence 117 AA;
XX
XX Query Match 68.6%; Score 615; DB 5; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 5.9e-51;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 58 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSK 117
XX Db 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSK 60
XX
XX QY 118 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
XX Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117
XX
XX RESULT 12
XX ADC45149
XX ID ADC45149 standard; protein; 117 AA.
XX
XX AC ADC45149;
XX
XX DT 18-DEC-2003 (first entry)
XX
XX DE S. pneumoniae antigenic protein SP043.
XX
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX
XX OS Streptococcus pneumoniae.
XX
XX XX US6573082-B1.
XX
XX PD 03-JUN-2003.
XX
XX PF 28-MAR-2000; 2000US-00536784.
XX
XX PR 31-OCT-1996; 96US-0029960P.
```

```
PR 30-OCT-1997; 97US-00961083.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX
XX WPI; 2003-764574/72.
XX N-PSDB; ADC45148.
XX
XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX useful for producing vaccines for prevention or attenuation of infection
XX by Streptococcus pneumoniae.
XX
XX Example 1; SEQ ID NO 68; 58pp; English.
XX
XX The invention relates to an isolated polynucleotide consisting of a
XX Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding the
XX SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX antigens. Also included are making a recombinant vector by inserting the
XX nucleic acid into a vector, an isolated polynucleotide consisting of at
XX least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX acids are useful as DNA vaccine against Streptococcus pneumoniae
XX infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX antigen nucleic acids are useful as probes for use in diagnostic methods
XX for detecting S. pneumoniae gene expression. The present sequence
XX represents an S. pneumoniae antigenic protein.
XX
XX Sequence 117 AA;
XX
XX Query Match 68.6%; Score 615; DB 7; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 5.9e-51;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 58 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSK 117
XX Db 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKKEENKPTFDVSK 60
XX
XX QY 118 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
XX Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117
XX
XX RESULT 13
XX ABP39023
XX ID ABP39023 standard; protein; 778 AA.
XX
XX AC ABP39023;
XX
XX DT 24-JUL-2002 (first entry)
XX
XX DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.
XX
XX KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
XX OS Staphylococcus epidermidis.
XX
XX PN US6380370-B1.
XX
XX PD 30-APR-2002.
XX
XX PF 13-AUG-1998; 98US-00134001.
XX
XX PR 14-AUG-1997; 97US-0055779P.
XX PR 08-NOV-1997; 97US-0064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
```

```

DR N-PSDB; ABN91568.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermis
PT polypeptide, useful for diagnosing and treating bacterial infections.
XX
PS Disclosure; SEQ ID NO 3868; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP5124 to ABP3960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences can
CC also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life cycle
CC or inhibit S. epidermidis infection. N.B. The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from the USPTO web site
XX
SQ Sequence 778 AA;
Query Match 13.5%; Score 121; DB 5; Length 778;
Best Local Similarity 26.8%; Pred. No. 0.022;
Matches 51; Conservative 26; Mismatches 65; Indels 48; Gaps 10;
QY 6 DPAR--NTTVKGFILNKDTGEVSELKPHRVTVTTONGKEMSTIVSEEDFILPVYK---- 59
Db 601 DFSKVPNVQGDVEQKAEDSVNAQSLKP-----ITINGKQIKQSVKSGTKVLPKSHKVM 656
QY 60 --GELEKGYQFDGW---EISGFE-----GKKDAGYVIN--LSKDTFIKPVFKKIEKK 105
Db 657 TDGELTMP-DMTGWTKEVDLAFEDLTIKIVSTKGNFGVFNQISKGQIIK----- 705
QY 106 EENKPTFDVS-----KKKDNPNVNSQLNESHKEDLQREHHSOKSDSKDVTATVLDKN 161
Db 706 ---NKDKIEVLSAEDTDDQEKTDDESDNKSKKDADEHSTSSSTKN-----DKS 756
QY 162 NISSKSTNN 171
Db 757 NADSKNDSDD 766
XX
RESULT 14
AD506368
ID ADS06368 standard; protein; 778 AA.
XX
AC ADS06368;
XX
DT 04-NOV-2004 (first entry)
XX
DE Staphylococcus epidermis polypeptide seqid 5663.
XX
KW antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
KW recombinant expression vector; infection; computer readable medium;
KW computer based system.
XX
OS Staphylococcus epidermidis.
XX
PN US2004147734-A1.
XX
PD 29-JUL-2004.
XX
PF 01-DEC-2003; 2003US-00724972.
XX
PR 08-NOV-1997; 97US-0064964P.
PR 13-AUG-1998; 98US-00134001.
PR 29-NOV-1999; 99US-00450969.
XX
PA (DOUC/) DOUCETTE-STAMM L.
PA (BUSH/) BUSH D.
XX
PI Doucette-Stamm L, Bush D;
XX
DR WPI; 2004-580138/56.

```

```

DR N-PSDB; ADS02596.
XX
PT New isolated polypeptide and encoding nucleic acid derived from
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT treating an S. epidermidis bacterial infection.
XX
PS Claim 17; SEQ ID NO 5663; 741pp; English.
XX
CC The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC given in the specification. Also described are: a recombinant expression
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an S. epidermidis polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for S. epidermidis infection; a recombinant or substantially
CC pure preparation of an S. epidermidis polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection; detecting the presence of a Staphylococcus nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the Staphylococcus genome of
CC commercial importance; a computer based system for identifying fragments
CC of the Staphylococcus plasmids of commercial importance; identifying
CC commercially important nucleic acid fragments of the Staphylococcus
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the Staphylococcus genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
CC infection. This is the amino acid sequence of a S. epidermis protein of
CC the invention.
XX
SQ Sequence 778 AA;
Query Match 13.5%; Score 121; DB 8; Length 778;
Best Local Similarity 26.8%; Pred. No. 0.022;
Matches 51; Conservative 26; Mismatches 65; Indels 48; Gaps 10;
QY 6 DPAR--NTTVKGFILNKDTGEVSELKPHRVTVTTONGKEMSTIVSEEDFILPVYK---- 59
Db 601 DFSKVPNVQGDVEQKAEDSVNAQSLKP-----ITINGKQIKQSVKSGTKVLPKSHKVM 656
QY 60 --GELEKGYQFDGW---EISGFE-----GKKDAGYVIN--LSKDTFIKPVFKKIEKK 105
Db 657 TDGELTMP-DMTGWTKEVDLAFEDLTIKIVSTKGNFGVFNQISKGQIIK----- 705
QY 106 EENKPTFDVS-----KKKDNPNVNSQLNESHKEDLQREHHSOKSDSKDVTATVLDKN 161
Db 706 ---NKDKIEVLSAEDTDDQEKTDDESDNKSKKDADEHSTSSSTKN-----DKS 756
QY 162 NISSKSTNN 171
Db 757 NADSKNDSDD 766
XX
RESULT 15
AAG81779
ID AAG81779 standard; protein; 746 AA.
XX
AC AAG81779;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
KW endocarditis.
XX
OS Staphylococcus epidermidis.
XX

```

PN W0200134809-A2.  
XX 17-MAY-2001.  
XX 09-NOV-2000; 2000WO-US030782.  
XX 09-NOV-1999; 99US-0164258P.  
XX (GLAX ) GLAXO GROUP LTD.  
XX Kimmerly WJ;  
XX WPI; 2001-316495/33.  
XX N-PSDB; AAH52629.  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis.  
XX Claim 18; Page 208; 2188pp; English.  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX Sequence 746 AA;  
SQ  
Query Match 13.3%; Score 119; DB 4; Length 746;  
Best Local Similarity 27.0%; Pred. No. 0.032;  
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
QY 20 KDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQPDGW-- 71  
DB 585 EDSVNAQSLKP-----ITINGKQIKQOQSVKSGTKVLPKSHKVMMLMTDGLTMP-DMTGWTK 639  
QY 72 -EISGPE-----GKKDAGYVIN--LSKDTPIKVPFKKIEKKKEENKPTFDVS---K 117  
DB 640 EDVLAPEDTLKLKVSFKNGFVTNQISIKQIILK-----NKKIEVLSAED 686  
QY 118 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKQVTVATVLDKNNISSKSTNN 171  
DB 687 TDDQDEKTDSDSDNKSCKKADHDHSNTSSSTKN-----DKSNADSKNDSDD 734  
RESULT 16  
ADZ72253  
ID ADZ72253 standard; protein; 354 AA.  
XX ADZ72253;  
AC ADZ72253;  
XX 14-JUL-2005 (first entry)  
XX Plasmodium falciparum MSP-3-like protein, MSP-3-1 SEQ ID NO: 2.  
DE Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;  
KW infection; merozoite surface protein 3-like protein; MSP-3-1; antigen.  
XX

OS Plasmodium falciparum.  
XX EP1526178-A1.  
XX 27-APR-2005.  
PD 24-OCT-2003; 2003EP-00292673.  
XX 24-OCT-2003; 2003EP-00292673.  
XX (INSP ) INST PASTEUR.  
PA Drulhe P;  
PI WPI; 2005-323987/34.  
XX N-PSDB; ADZ72252.  
DR Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
XX falciparum, which encode proteins useful for preparing vaccine  
PT compositions against malaria.  
PT Disclosure; SEQ ID NO 2; 137pp; English.  
PS The present invention relates to the protection against malaria. More  
XX particularly, the invention pertains to a family of MSP-3 (merozoite  
CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
CC falciparum, highly conserved in P. falciparum strains, simultaneously  
CC expressed in P. falciparum at the erythrocytic stages and encoding  
CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at  
CC their N-terminal extremity and which are located at the merozoite  
CC surface. The characterization of this gene family enables the definition  
CC of immunogenic and vaccine compositions against P. falciparum. The  
CC present sequence is the P. falciparum MSP-3-1 protein.  
XX Sequence 354 AA;  
SQ  
Query Match 12.5%; Score 112.5; DB 9; Length 354;  
Best Local Similarity 22.4%; Pred. No. 0.048;  
Matches 41; Conservative 35; Mismatches 62; Indels 45; Gaps 7;  
QY 10 NTTVGEFILN-KDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQF 69  
DB 153 STTKTEYAFKAKNAYEKAKNAYQKANQAVLKAKBASS-----YDYIL----- 194  
QY 69 DGWEISGR--EKKDAG-----YVINLSKDTPIKVPFKKIEKKKEE-----E 108  
DB 195 -GWEEGGGVPEKKKENMLSHLYSSKKNENISKENDVDLDKKEBAAETEEHELEEKNE 253  
QY 109 NKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKQVTVATVLDKNNISSKST 168  
DB 254 BETEISEISEDEEBEEREBEENKKKEQKQENNNQKDMEA-----QNLISKQ 308  
QY 169 TNN 171  
DB 309 NNN 311  
RESULT 17  
ABU25018  
ID ABU25018 standard; protein; 707 AA.  
XX AC ABU25018;  
XX DT 19-JUN-2003 (first entry)  
XX Protein encoded by Prokaryotic essential gene #10545.  
DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
KW Clostridium difficile.  
XX WO200277183-A2.





```
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 775 AA;
Query Match 12.4%; Score 111.5; DB 6; Length 775;
Best Local Similarity 24.9%; Pred. No. 0.18;
Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;
QY 20 KDTGEVSELPKPHRVTVTTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQPDGW-- 71
DB 614 EDSVNAQSLKP-----ITINGKQIKQOSVKSQTKVLPKSHKVLMTDGBELTMP-DMTGWTK 668
QY 72 -EISGPE-----GKDDAGYVIN--LSKDTTFIKPVFK-----KISEKKEENKPTF 113
DB 669 EDVLAFEDLTIKIKVSTKGGFVTTNQISKGQIIKNDKIEVLSABEDTDDQKTDSDSS 728
QY 114 DVSKKDNQPNVHSHKEDLQREHESQKSDSTKDVATVLDKNN 162
DB 729 DKSKDKVDKEDNSNASSSKNEKSNADSKNDSDDSTNETSGS--ERNN 775
RESULT 19
ADZ79635
ID ADZ79635 standard; protein; 647 AA.
XX AC ADZ79635;
XX DT 14-JUL-2005 (first entry)
XX DE P. falciparum GLURP-MSP3 fusion protein.
XX KW immune stimulation; fusion protein; glutamate-rich protein; GLURP;
XX KW merozoite surface protein 3; MSP3; immunotherapy; malaria; antimalarial;
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX PN WO2005040206-A1.
XX PD 06-MAY-2005.
XX PF 22-OCT-2004; 2004WO-EP012910.
XX PR 24-OCT-2003; 2003US-00691672.
XX PA (INSP ) INST PASTEUR.
XX PI Druilhe P;
XX WPI; 2005-355821/36.
XX N-PSDB; ADZ79636.
XX Chimeric molecule useful for preparing vaccine composition against
XX malaria, comprises glutamate-rich protein GLURP and Merozoite surface
XX protein 3 MSP3 moieties, and raises antibodies against moieties in mice
XX immunized with molecule.
XX Disclosure; SEQ ID NO 3; 79pp; English.
XX The invention relates to a chimeric molecule that comprises a glutamate-
XX rich protein (GLURP) moiety consisting of a polypeptide fragment (amino
XX acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite
XX surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380
XX of MSP3 (given as SEQ ID No:2) wherein the chimeric molecule raises
XX antibodies against both polypeptides in mice immunized with it. Also
XX described are: (i) a conjugate comprising the chimeric molecule of the
XX invention bound to a solid support, (ii) an immunogenic composition
XX comprising the chimeric molecule, the conjugate described above, or a
XX mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine
XX against malaria comprising the chimeric molecule, the conjugate described
XX above, or a mixture of GLURP and MSP3 antigens as an immunogen, in
XX association with a suitable vehicle, (iv) use of purified and/or
CC
```

```
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a
CC medicament against malaria, and (v) a medicament for passive
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP
CC and MSP3 antigens are useful for the preparation of a vaccine composition
CC against malaria. This sequence represents Plasmodium falciparum GLURP(27-
XX 500)-MSP3(212-380) fusion protein.
SQ Sequence 647 AA;
Query Match 12.3%; Score 110.5; DB 9; Length 647;
Best Local Similarity 22.8%; Pred. No. 0.17;
Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;
QY 14 KEFILNKDTGKRVSELPKPHRVTVTTIQNGKEM-----SSTIVSEEDFILPVYKGELE 63
DB 418 EEAVSEKNAHETVE---HEETVQSSENPKEKADNGVSNNSNELNENEFV-----ESE 468
QY 64 K-----GYQFD-GWEISGF--EGKKDAG-----YVINLSKDTTFIKPVFKIE 102
DB 469 KSEHEARSKAKEASSYDYILGWFGGVPPEHKKEENMLSHLYVSSKDKENISKENDVDLD 528
QY 103 EKKEENKPTFDVSKKKNPQVNHSLN-----ESHRKEDLQREHESQKSDS 149
DB 529 E-KEEABETEEELKEENKNEBETSEISEDEEBEKEENKKEKQEKQSQSNENNDQ 587
QY 150 TKDVTATVLDKNNISKSTTNN 171
DB 588 KKMEA-----QNLISKNNQNN 604
RESULT 20
AD019012
ID AD019012 standard; protein; 651 AA.
XX AC AD019012;
XX DT 12-AUG-2004 (first entry)
XX DE Amino acid sequence for P. falciparum GLURP-MSP3 hybrid.
XX KW Glutamate-rich protein; GLURP-MSP3 fusion protein;
XX KW merozoite surface protein 3; malarial vaccine; malaria; immune response;
XX KW antimalarial; immunostimulant.
XX OS Plasmodium falciparum.
XX OS Synthetic.
XX PN WO2004043488-A1.
XX PD 27-MAY-2004.
XX PF 06-NOV-2003; 2003WO-DK000759.
XX PR 12-NOV-2002; 2002DK-00001741.
XX PR 11-SEP-2003; 2003DK-00001307.
XX PA (STAT-) STATENS SERUM INST.
XX PI Theisen M, Jepsen S;
XX WPI; 2004-411650/38.
XX New antigen based vaccine comprising a fusion protein derived from
XX Plasmodium falciparum Glutamate-rich protein, useful in treating or
XX preventing malaria.
XX Disclosure; Fig 2C; 52pp; English.
XX The present invention relates to a fusion protein comprising Plasmodium
XX falciparum glutamate-rich protein (GLURP) coupled to P. falciparum
XX merozoite surface protein 3 (MSP3). The GLURP-MSP3 fusion protein is
XX useful as an antigen based vaccine against malaria. Also disclosed is the
```

CC polynucleotide sequence encoding the GLURP-MSP3 fusion protein. The  
CC polynucleotide sequence is also useful in preparing a vaccine. The  
CC vaccine is useful in treating and preventing malaria and for inducing an  
CC immune response against malaria. The present sequence represents P.  
CC falciparum GLURP-MSP3 hybrid.  
XX  
SQ Sequence 651 AA;

Query Match 12.3%; Score 110.5; DB 8; Length 651;  
Best Local Similarity 22.8%; Pred. No. 0.17; Mismatches 59; Gaps 10;  
Matches 46; Conservative 38; Indels 59; Indels 59; Gaps 10;  
QY 14 KEFILNKDTGSEVSELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 63  
DB 422 EBAVSEKNAHETVE---HEETVSQESNPEKADNGVNSQNSNNELNEFEV-----ESE 472  
QY 64 K-----GYQFD-GWEISGF--EGKKDAG-----YVINLSKDTFIKVPVKIE 102  
DB 473 KSEHARSKAKEASSYDILGWFEFGGVPEHKKEENMLSHLYVSSKDKENISKENDVDLD 532  
QY 103 EKKEENKPTFDVSKKDNPNQVNHSQLN-----ESHRKEDLQREHSQKSDS 149  
DB 533 E-KEEABETEEELKEEETSEISEDEEEEEKEEENKKEKQEQKQSNENNDQ 591  
QY 150 TKDVTATVLDKNNISKSTTNN 171  
DB 592 KQMEA-----QNLISKNNNN 608

RESULT 21  
AD019010  
ID AD019010 standard; protein; 651 AA.  
XX  
AC AD019010;  
XX  
DT 12-AUG-2004 (first entry)  
XX  
DE P. falciparum GLURP-MSP3 fusion protein.  
XX  
KW Glutamate-rich protein; GLURP-MSP3 fusion protein;  
KW merozoite surface protein 3; malarial vaccine; malaria; immune response;  
KW antimalarial; immunostimulant.  
XX  
OS Plasmodium falciparum.  
OS Synthetic.  
XX  
PN WO2004043488-A1.  
XX  
PD 27-MAY-2004.  
XX  
PF 06-NOV-2003; 2003WO-DK000759.  
XX  
PR 12-NOV-2002; 2002DK-00001741.  
PR 11-SEP-2003; 2003DK-00001307.  
XX  
PA (STAT-) STATENS SERUM INST.  
XX  
PI Theisen M, Jepsen S;  
XX  
DR WPI; 2004-411650/38.  
DR N-PSDB; AD019011.  
XX  
XX New antigen based vaccine comprising a fusion protein derived from  
PT Plasmodium falciparum Glutamate-rich protein, useful in treating or  
PT preventing malaria.  
XX  
XX Claim 5; SEQ ID NO 1; 52pp; English.  
PS  
CC The present invention relates to a fusion protein comprising Plasmodium  
CC falciparum glutamate-rich protein (GLURP) coupled to P. falciparum  
CC merozoite surface protein 3 (MSP3). The GLURP-MSP3 fusion protein is  
CC useful as an antigen based vaccine against malaria. Also disclosed is the  
CC polynucleotide sequence encoding the GLURP-MSP3 fusion protein. The

CC polynucleotide sequence is also useful in preparing a vaccine. The  
CC vaccine is useful in treating and preventing malaria and for inducing an  
CC immune response against malaria. The present sequence represents P.  
CC falciparum GLURP-MSP3 fusion protein.  
XX  
SQ Sequence 651 AA;

Query Match 12.3%; Score 110.5; DB 8; Length 651;  
Best Local Similarity 22.8%; Pred. No. 0.17; Mismatches 59; Gaps 10;  
Matches 46; Conservative 38; Indels 59; Indels 59; Gaps 10;  
QY 14 KEFILNKDTGSEVSELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 63  
DB 422 EBAVSEKNAHETVE---HEETVSQESNPEKADNGVNSQNSNNELNEFEV-----ESE 472  
QY 64 K-----GYQFD-GWEISGF--EGKKDAG-----YVINLSKDTFIKVPVKIE 102  
DB 473 KSEHARSKAKEASSYDILGWFEFGGVPEHKKEENMLSHLYVSSKDKENISKENDVDLD 532  
QY 103 EKKEENKPTFDVSKKDNPNQVNHSQLN-----ESHRKEDLQREHSQKSDS 149  
DB 533 E-KEEABETEEELKEEETSEISEDEEEEEKEEENKKEKQEQKQSNENNDQ 591  
QY 150 TKDVTATVLDKNNISKSTTNN 171  
DB 592 KQMEA-----QNLISKNNNN 608

RESULT 22  
AAB18278  
ID AAB18278 standard; protein; 665 AA.  
XX  
AC AAB18278;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:135.  
XX  
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
KW antimalarial; malaria; protozoicide; infection; insecticide.  
XX  
OS Plasmodium falciparum.  
XX  
PN WO2000025728-A2.  
XX  
PD 11-MAY-2000.  
XX  
PF 05-NOV-1999; 99WO-US026796.  
XX  
PR 05-NOV-1998; 98US-0107131P.  
XX  
PA (HOFF/) HOFFMAN S.  
PA (CARU/) CARUCCI D.  
PA (GARD/) GARDNER M.  
PA (VENT/) VENTER J C.  
XX  
PI Hoffman S, Carucci D, Gardner M, Venter JC;  
XX  
DR WPI; 2000-365347/31.  
XX  
XX Proteins encoded by chromosome 2 of the human malarial parasite,  
PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection.  
XX  
PS Disclosure; Page 321-322; 577pp; English.  
XX  
CC The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II). (I) and  
CC (II) are useful for the development of vaccines against P. falciparum  
CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to  
CC immunogens comprising the sequences of (I), are useful in the detection





XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
XX EF1033405-A2.  
XX 06-SRP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 21-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.



Qy	9	RNTTVK--BFLNKTGVESELKP-----HRVTVTIQNG-----KEMSSITVSE 50
Db	614	RNTKYKSDYLLDNKKGSKFKRNTSYVLBSPLHLIGDIVDNNIKKKGKKGKIKTIVSD 673
Qy	51	EDFILPVYKGELEKGQPGWRIIG---PEGKQAGYVINLSKDTFTKPVPKIEKKGE 107
Db	674	DMTSPWNIEKYNNEQEKKGIVGNLSYDTKIPPPKIEKGRKIKK--KKIEKKKK 731
Qy	108	ENK-----PTF-----DVSKKKQNPQVNHSQL----- 129
Db	732	EKKENNNFLYNDYSSYPKYGDNNENPVIKYIEREKQFKCFDHPNFPKFLPHYN 791
Qy	130	---NEGHRK---EDLQREHSSQKSSSTKD-VTATVLDKKNISSSKSTNNPNK 174
Db	792	PMKNKKNKKNKNKKNVARNYPNYNTSSSKGYSYVNFSLSDSLPSSNBYESSNBE 843

**RESULT 28**

RESOL 28  
ABU24404  
ID ABU24404 standard: protein: 903 AA.

XX  
AC

DT 19-JUN-2003 (first entry)

protein encoded by prokaryotic essential gene #9931

**KW** Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS *Clostridium botulinum*.

PN WO200277183-A2.

PD 03-OCT-2002.

21-MAR-2002: 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 25-OCT-2001; 2001US-0342923P.

PR 06-MAR-2002; 2002US-0362699P.

PA (BLIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C

1000

DR N-PSDB; ACA28274.

PT New antisense nuc

PT isolate candidate

PS Claim 25; SEQ ID

CC The invention relates to an isolated nucleic acid

cells also produced sequences giving rise to the proliferation of a cell. Also included are:  
 (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an

PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
PT immunized with molecule.  
XX  
XX Claim 2; SEQ ID NO 2; 79pp; English.  
XX  
CC The invention relates to a chimeric molecule that comprises a glutamate-  
CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
CC acid residues 25-514) of GLURP (given as SEQ ID NO.1) and a Merozoite  
CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
CC of MSP3 (given as SEQ ID NO.2), wherein the chimeric molecule raises  
CC antibodies against both polypeptides in mice immunized with it. Also  
CC described are: (i) a conjugate comprising the chimeric molecule of the  
CC invention bound to a solid support, (ii) an immunogenic composition  
CC comprising the chimeric molecule, the conjugate described above, or a  
CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
CC against malaria comprising the chimeric molecule, the conjugate described  
CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
CC association with a suitable vehicle, (iv) use of purified and/or  
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
CC medicament against malaria, and (v) a medicament for passive  
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents Plasmodium falciparum MSP3  
CC protein (amino acid residues 212-380).  
XX  
XX Sequence 169 AA;

Query Match 11.5%; Score 103.5; DB 9; Length 169;  
Best Local Similarity 25.2%; Pred. No. 0.13;  
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;  
QY 41 KEMSSIVSEEPFILPVYKGELEKGYQFGWEISGF--EGKDGAG-----YVINLSKDTF 93  
DB 1 KEASS-----YDYIL-----GWBFGGVPEHKKENMLSHLYVSSKDKENI 41  
QY 94 IKPVFKKIBKKKEENKPTFVSKKDNQVNHSQLN-----ESHKEDLQR 140  
DB 42 SKENDVDLDE-KEEEAEETEEBELEKNEETEESEISEDEEBEEEEKEEKKKQBK 100  
QY 141 BEHSQKSDSTKQVATVLDKNNISSKSTNN 171  
DB 101 EQSNENNDDQKDEA-----QNLISKNNNN 126

RESULT 30  
ID ADW88441 standard; protein; 645 AA.

XX ADW88441;

XX 21-APR-2005 (first entry)

XX Staphylococcus aureus hybrid polypeptide 0657nHybrid3.

XX ORF0657n; vaccine; antibacterial; protein engineering;

XX Staphylococcus aureus infection; muten.

XX Staphylococcus aureus.

OS Synthetic.

XX WO2005009378-A2.

XX 03-FEB-2005.

XX 22-JUL-2004; 2004WO-US023522.

XX 24-JUL-2003; 2003US-0469840P.

XX (MERI ) MERCK & CO INC.

XX Anderson AS, Kuklin N, Jansen KU;

XX WPI; 2005-123069/13.  
XX  
PT Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
PT useful for inducing protective immune response in humans against  
PT Staphylococcus aureus infection.  
XX  
XX Claim 7; SEQ ID NO 10; 84pp; English.

XX The present sequence is that of a Staphylococcus aureus protein ORF0657n  
CC hybrid polypeptide denoted 0657nHybrid3. This is an example of claimed  
CC hybrid polypeptide immunogens of the invention ADW8439-ADW8474 that  
CC comprise a modified S. aureus ORF0657n sequence ADW8433-ADW8438  
CC containing amino acid substitutions that increase sequence similarity to  
CC ORF0190 ADW8432. The hybrid polypeptides contain one or more epitopes  
CC for ORF0657n and ORF0190. They were designed by taking into account the  
CC similarity and differences between native ORF0657n and ORF0190 protein  
CC sequences. The invention also provides nucleic acids encoding these  
CC hybrid polypeptides, and a method for evaluating the ability of an  
CC immunogen to produce a protective immune response against Staphylococcus  
CC infection using an animal (mouse or rat) model. The hybrid polypeptides  
CC having therapeutic and diagnostic applications, such as being used to  
CC provide protective immunity against S. aureus infection, being used to  
CC generate antibodies to detect the presence of S. aureus, and being used  
CC to generate therapeutic antibodies that target S. aureus.

XX Sequence 645 AA;

Query Match 11.4%; Score 102; DB 9; Length 645;  
Best Local Similarity 24.1%; Pred. No. 1.1;  
Matches 46; Conservative 33; Mismatches 68; Indels 44; Gaps 8;

QY 1 KIVVKDFARNTTVKFEILNKDTGVESELKPHRVTVTIQNGKEMSTIV-----SEEDF 53  
DB 380 KYVMETNDYWKDFI-----VEGQVRVTVKDKNNGRTLPIPIEGTKTVYDA 429  
QY 54 ILPVYKGELEKGYQFDGWEISGFEGKDGAGYVINLSKDTFIKPVFKKIEEKEENKPTF 113  
DB 430 IVKHVKTID-----YEGQY---HVRIDKDAFTKANTDKSNKKEQDQNSAKK 474  
QY 114 DV-----SKKDNQVNHSQLNESHKEDLQ-----REEHSQKSDSTKQVT-ATVLDKNNI 163  
DB 475 EATPATPSKTPSPVKESESQKDSQKDNKQLPSVEKENDASSBSGDKDKTPATKPTKGEV 534

QY 164 SSKSTNNPNK 174  
DB 535 ESSSTT--PTK 543

RESULT 31  
ID ABB61977 standard; protein; 564 AA.

XX ABB61977;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 12723.

XX Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX











CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR284074A1, which  
 CC contains only 2344 sequences.

XX SQ Sequence 643 AA;

Query Match 11.1%; Score 99.5; DB 8; Length 643;  
 Best Local Similarity 23.8%; Pred. No. 1.9;  
 Matches 44; Conservative 39; Mismatches 67; Indels 35; Gaps 10;  
 QY 3 VKRD--FARNVTTKFIFLNKDTGVSELSKPHRVTVTIQNGKEMSSIVSEEDPILP--VY 58  
 Db 293 IAKDKLFSLESLKEY-----GKEKVNYSRLP-----NTEPLTSVLEKNKFLVNPY 342  
 QY 59 KGEI-----BKQVQPD-----GWEISGFPGKKDAGVIVNLSKDTPIKPVFKKIEKKEE 108  
 Db 343 KKKLILREEDKYSFEDDEBEFGNELLSYNKLKNEVLVFNITTTILKP-----FEQKKIVE 398  
 QY 109 NKPTPDVSKKONPQVNHQSLNESHKEDLQREHS--QKSDTKDVTATVLDKN--NLS 164  
 Db 399 D----FNPYNLNDNLKIKRLNGSKQKVEQETKSPTPQKTVKQETQKVGNTQRE 455  
 QY 165 SKSTT 169  
 Db 456 KKSET 460

RESULT 39  
 ABP5413  
 ID ABP5413 standard; protein; 1384 AA.  
 AC ABP5413;  
 XX  
 DT 04-FEB-2003 (first entry)  
 XX Human MDDT-22 protein SEQ ID NO:22.

DE Human MDDT-22 protein SEQ ID NO:22.  
 XX Human; MDDT; molecules for disease detection and treatment; anti-HIV;  
 KW antiallergic; antinflammatory; antianaemic; antiparkinsonian; nootropic;  
 KW anticonvulsant; antinfertility; antiarteriosclerotic; antiasthmatic;  
 KW immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;  
 KW antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;  
 KW osteopathic; antiarthritic; antiparasitic; antihelminthic; antipsoriatic;  
 KW uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
 KW viricide; protozoacide; fungicide; gene therapy; cell proliferative;  
 KW cancer; developmental disorder; neurological disorder; infection;  
 KW reproductiv disorder; autoimmune disorder; inflammatory disorder.

XX Homo sapiens.  
 OS  
 XX WO200278420-A2.  
 PN 10-OCT-2002.  
 PD  
 XX 29-MAR-2002; 2002WO-US009809.  
 PF  
 XX 30-MAR-2001; 2001US-0280387P.  
 PR 05-APR-2001; 2001US-0282335P.  
 PR 13-APR-2001; 2001US-0283663P.  
 PR 19-APR-2001; 2001US-0285484P.  
 PR 18-JAN-2002; 2002US-0350702P.  
 PR 25-JAN-2002; 2002US-0351749P.

XX (INCY-) INCYTE GENOMICS INC.  
 XX Lu DAM, Arvizu CS, Gandhi AR, Hafalia AJA, Ding L, Lu Y;  
 PI Ramkumar J, Swannakar A, Tang YT, Yue H, Tran B, Lee SY, Warren BA;  
 PI Nguyen DB, Thangavelu K, Yao MG, Elliott VS, Baughn ME, Emerling BM;

PI Lal PG, Gietzen KJ, Becha SD, Marquis JP, Kable AE;  
 XX WPI; 2003-058385/05.  
 DR N-PSDB; ABQ83880.  
 XX  
 XX New human molecules for disease detection and treatment, useful for  
 PT diagnosing, treating or preventing autoimmune or inflammatory disorders  
 PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
 PT cancer or hepatitis.  
 XX  
 PS Claim 1; Page 207-211; 238pp; English.

XX The present invention describes 23 human molecules for disease detection  
 CC and treatment (MDDT-1 to 23) (see ABP55392 to ABP5414). The human MDDT-1  
 CC to 23 proteins (I) are encoded by the sequences given in ABQ83889 to  
 CC ABQ83891. (I) can have various activities depending on the cells and  
 CC tissues in which they are expressed. These activities include: anti-HIV;  
 CC antiallergic; antinflammatory; antianaemic; antiparkinsonian; nootropic;  
 CC anticonvulsant; antinfertility; antiarteriosclerotic; antiasthmatic;  
 CC immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;  
 CC antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;  
 CC osteopathic; antiarthritic; antiparasitic; antihelminthic; antipsoriatic;  
 CC uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
 CC viricide; protozoacide; and fungicide. (I) and the polynucleotides  
 CC encoding them can be used in gene therapy. (I), polynucleotides, agonists  
 CC and antagonists from the present invention can be used for diagnosing,  
 CC treating or preventing disorders associated with aberrant expression of  
 CC MDDT, particularly cell proliferative (e.g. cancer), developmental  
 CC disorders, neurological disorders, reproductive disorders, or  
 CC autoimmune/inflammatory disorders, or viral, bacterial, fungal,  
 CC parasitic, protozoal or helminthic infections. They are also useful in  
 CC the assessment of the effects of exogenous compounds on the expression of  
 CC nucleic acid and amino acid sequences of proteins associated with MDDT  
 XX  
 SQ Sequence 1384 AA;

Query Match 11.1%; Score 99.5; DB 6; Length 1384;  
 Best Local Similarity 21.7%; Pred. No. 5.5;  
 Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;  
 QY 14 KEFILLKDTGVSELSKPHRVTVTIQNGKEMSSIVSEEDPILPVYKGELEKGYQF----- 68  
 Db 1013 KEFILLQNEQIEISLQK-KEIERTQQRKEMSVKESQYIATQYKKAIDLDGQELRLTRE 1071  
 QY 69 -----DGWEISGFEGKKDA-----GVVINLSK----- 90  
 Db 1072 QVQNSHTLAEARHQVQAQREIELSSELEDMKQLSKEKDAHGHNLAELGASKVREAH 1131  
 QY 91 -----DTPFKPVFKKIEKKE-----BENKPTPDVSKKONPQVNHQSLNESHKRE 136  
 Db 1132 LEARMQASIKKLAEVESLKEAYHMEMISHQENHAKWKIS--ADSKSSVQQLNQLLEKA 1189  
 QY 137 DLQREHSQKSDSTKDTATVLDKN 163  
 Db 1190 KLELEE---AQDTVSNLHQVQDRNEV 1213

RESULT 40  
 ABP55393  
 ID ABP55393 standard; protein; 1404 AA.  
 XX  
 AC ABP55393;  
 XX  
 DT 04-FEB-2003 (first entry)  
 XX  
 DE Human MDDT-2 protein SEQ ID NO:2.

XX Human; MDDT; molecules for disease detection and treatment; anti-HIV;  
 KW antiallergic; antinflammatory; antianaemic; antiparkinsonian; nootropic;  
 KW anticonvulsant; antinfertility; antiarteriosclerotic; antiasthmatic;  
 KW immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;  
 KW antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;  
 KW osteopathic; antiarthritic; antiparasitic; antihelminthic; antipsoriatic;

KW uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
KW virucide; protozoacide; fungicide; gene therapy; cell proliferative;  
KW cancer; developmental disorder; neurological disorder; infection;  
KW reproductive disorder; autoimmune disorder; inflammatory disorder.

OS Homo sapiens.

PN WO200278420-A2.

XX 10-OCT-2002.

XX 29-MAR-2002; 2002WO-US009809.

XX 30-MAR-2001; 2001US-0280387P.

PR 05-APR-2001; 2001US-0282335P.

PR 13-APR-2001; 2001US-0283663P.

PR 19-APR-2001; 2001US-0285484P.

PR 18-JAN-2002; 2002US-0350702P.

XX 25-JAN-2002; 2002US-0351749P.

XX (INCY-) INCYTE GENOMICS INC.

PI Lu DAM, Arvizu CS, Gandhi AR, Hafalia AJA, Ding L, Lu Y;

PI Ramkumar J, Swarnakar A, Tang YT, Yue H, Tran B, Lee SY, Warren BA;

PI Nguyen DB, Thangavelu K, Yao MG, Elliott VS, Baughn MR, Emerling BM;

PI Lal PG, Gietzen KJ, Becha SD, Marquis JP, Kable AE;

XX WPI; 2003-058385/05.

XX N-PSDB; ABQ83860.

XX New human molecules for disease detection and treatment, useful for  
XX diagnosing, treating or preventing autoimmune or inflammatory disorders  
XX (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
XX cancer or hepatitis.

XX Claim 1; Page 172-175; 238pp; English.

XX The present invention describes 23 human molecules for disease detection  
XX and treatment (MDT-1 to 23) (see ABP55392 to ABP55414). The human MDT-1  
XX to 23 proteins (I) are encoded by the sequences given in ABQ83859 to  
XX ABQ83881. (I) can have various activities depending on the cells and  
XX tissues in which they are expressed. These activities include: anti-HIV;  
XX anti-allergic; anti-inflammatory; antianaemic; antiparkinsonian; nootropic;  
XX anticonvulsant; antiinfertility; antiarteriosclerotic; antiasthmatic;  
XX immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;  
XX antidiabetic; nephrotropic; antigout; thyromimetic; neuroprotective;  
XX osteopathic; antiarthritic; antiparasitic; antihelminthic; antipeoriatic;  
XX uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
XX virucide; protozoacide; and fungicide. (I) and the polynucleotides  
XX encoding them can be used in gene therapy. (I), polynucleotides, agonists  
XX and antagonists from the present invention can be used for diagnosing, or  
XX treating or preventing disorders associated with aberrant expression of  
XX MDDT, particularly cell proliferative (e.g. cancer), developmental  
XX disorders, neurological disorders, reproductive disorders, or  
XX autoimmune/inflammatory disorders, or viral, bacterial, fungal,  
XX parasitic, protozoal or helminthic infections. They are also useful in  
XX the assessment of the effects of exogenous compounds on the expression of  
XX nucleic acid and amino acid sequences of proteins associated with MDDT

XX Sequence 1404 AA;

Query Match 11.1%; Score 99.5; DB 6; Length 1404;

Best Local Similarity 21.7%; Pred. No. 5.7;

Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;

QY 14 KEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILFVYKGELEKGYQF----- 68

DB 1033 KEFILNQCEISQLK-KEIERTQQRKEMESVMKEQEQYIATQYKEAIDLGLRLTRE 1091

QY 69 -----DGWEISGPEGKDA-----GYVNLK----- 90

DB 1092 QVQNSHTLAEARHQVQVQRETERLSLEEDMKQLSEKDAHGHNLAELGASKVREAH 1151

QY 91 -----DTPIKPVFKKIEEKE-----BENKPTFDVSKKDNQPVNHSQLNESHKKE 136  
DB 1152 LEARNQAEIKKLAEVESLKEAYHMEMISHQENHAKWKIS--ADSQKSSVQQLNEQLEKA 1209  
QY 137 DLQREHSQKSDSTKDVVTATVLDKNNI 163  
DB 1210 KLELEE---AQDTVSNLHQVQDRNEV 1233

Search completed: April 24, 2006, 14:50:30

Job time : 111.188 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 27.0537 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773  
Perfect score: 897  
Sequence: 1 KIVVKDFARNTTVKKEFILNK.....ATVLDKNNISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5\_COMB.pep.\*  
2: /cgn2\_6/prodata/1/1aa/6\_COMB.pep.\*  
3: /cgn2\_6/prodata/1/1aa/HCTUS\_COMB.pep.\*  
4: /cgn2\_6/prodata/1/1aa/PTUS\_COMB.pep.\*  
5: /cgn2\_6/prodata/1/1aa/RE\_COMB.pep.\*  
6: /cgn2\_6/prodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	773	2	US-09-590-991-8
2	894	99.7	637	2	US-09-107-433-3169
3	894	99.7	2138	2	US-09-583-110-5274
4	615	68.6	117	2	US-08-961-083-68
5	615	68.6	117	2	US-09-536-784-68
6	615	68.6	117	2	US-09-765-271-68
7	615	68.6	117	2	US-09-765-272A-68
8	121	13.5	778	2	US-09-134-001C-3868
9	119	13.3	746	2	US-09-710-279-652
10	104	11.6	347	2	US-09-248-796A-16224
11	97.5	10.9	2468	2	US-09-976-594-726
12	97.5	10.9	2468	2	US-09-538-092-1135
13	97.5	10.9	2522	2	US-09-949-016-10237
14	96	10.7	902	2	US-09-134-001C-3157
15	95	10.6	348	2	US-09-538-092-1316
16	94.5	10.5	277	2	US-09-830-230A-651
17	93.5	10.4	654	2	US-10-172-502-10
18	93	10.4	280	2	US-09-248-796A-17646
19	93	10.4	715	2	US-08-669-286-7
20	93	10.4	715	2	US-09-469-253-7
21	93	10.4	715	2	US-09-642-146-7
22	92.5	10.3	1702	2	US-08-296-791-5
23	92.5	10.3	1702	2	US-09-839-996-5
24	92.5	10.3	1702	2	US-10-080-505-5
25	92.5	10.3	1702	2	US-10-645-655-5
26	92.5	10.3	1702	4	PCT-US95-10661A-5
27	92	10.3	299	2	US-09-710-279-1888

28	92	10.3	309	2	US-09-134-001C-5667	Sequence 5667, Ap
29	92	10.3	561	1	US-08-360-673-2	Sequence 2, Appli
30	92	10.3	561	5	US-09-640-305-2	Sequence 2, Appli
31	91	10.1	442	2	US-09-134-001C-3033	Sequence 3033, Ap
32	90.5	10.1	743	2	US-08-910-925-3	Sequence 3, Appli
33	90.5	10.1	743	2	US-09-949-016-6261	Sequence 6261, Ap
34	90.5	10.1	758	2	US-09-949-016-8288	Sequence 8288, Ap
35	90	10.0	785	2	US-09-710-279-264	Sequence 264, App
36	89.5	10.0	262	2	US-09-248-796A-21451	Sequence 21451, A
37	89.5	10.0	402	2	US-09-464-483-4	Sequence 4, Appli
38	89.5	10.0	402	2	US-09-414-664-4	Sequence 4, Appli
39	89.5	10.0	511	2	US-09-198-452A-509	Sequence 509, App
40	89.5	10.0	511	2	US-09-438-185A-475	Sequence 475, App
41	89.5	10.0	529	2	US-09-464-483-2	Sequence 2, Appli
42	89.5	10.0	553	2	US-09-248-796A-16588	Sequence 16588, A
43	89.5	10.0	299	2	US-09-328-352-6567	Sequence 6567, Ap
44	89	9.9	299	2	US-09-134-001C-3148	Sequence 3148, Ap
45	88	9.8	443	2	US-09-134-001C-3148	Sequence 3148, Ap

ALIGNMENTS

RESULT 1

US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%;	Score 897;	DB 2;	Length 773;
Best Local Similarity	100.0%;	Pred. No. 8.9e-83;		
Matches 174;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTTIIONGKEMSTIVSEEDFILPVYKG	60	
Db	600	KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTTIIONGKEMSTIVSEEDFILPVYKG	659	
QY	61	ELEKGQFDGWEISGFEKGDAGVYINLSKDTFTKPVFKIEEKEENKPTFDVSKKD	120	
Db	660	ELEKGQFDGWEISGFEKGDAGVYINLSKDTFTKPVFKIEEKEENKPTFDVSKKD	719	
QY	121	NPQVNHSQLNESHKREDLQREHKSQSDSTKDVTATVLDKNNISSKSTNNPNK	174	
Db	720	NPQVNHSQLNESHKREDLQREHKSQSDSTKDVTATVLDKNNISSKSTNNPNK	773	

RESULT 2

US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...637
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-433-3169

Query Match          99.7%; Score 894; DB 2; Length 637;
Best Local Similarity 99.4%; Pred. No. 1.4e-82;
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDPARNTTVKEFIILNKDTGEVSELKPHRVTTVTIQNGKEMSTIVSEEDFILPVYKG 60
DB 430 KIVVKDPARNTTVKEFIILNKDTGEVSELKPHRVTTVTIQNGKEMSTIVSEEDFILPVYKG 489
QY 61 ELEKGYQFDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEEKKKEENKPTFDVSKKKD 120
DB 490 ELEKGYQFDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEEKKKEENKPTFDVSKKKD 549
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 174
DB 550 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 603

RESULT 3
US-09-583-110-5274
; Sequence 5274, Application US/09583110
; Patent No. 6699703
;
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
```

```
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5274
; LENGTH: 2138
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5274

Query Match          99.7%; Score 894; DB 2; Length 2138;
Best Local Similarity 99.4%; Pred. No. 7.3e-82;
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDPARNTTVKEFIILNKDTGEVSELKPHRVTTVTIQNGKEMSTIVSEEDFILPVYKG 60
DB 1931 KIVVKDPARNTTVKEFIILNKDTGEVSELKPHRVTTVTIQNGKEMSTIVSEEDFILPVYKG 1990
QY 61 ELEKGYQFDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEEKKKEENKPTFDVSKKKD 120
DB 1991 ELEKGYQFDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEEKKKEENKPTFDVSKKKD 2050
QY 121 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 174
DB 2051 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2104

RESULT 4
US-08-961-083-68
; Sequence 68, Application US/08961083
; Patent No. 6159469
;
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
;
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340F2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-083-68

Query Match          68.6%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.7e-55;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 YKGELEKGYQFDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIEEKKKEENKPTFDVSK 117
```



```
Db      1 YGGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 60
QY      118 KKNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 174
Db      61 KKNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 5
US-09-536-784-68
; Sequence 68, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match      68.6%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.7e-55;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      58 YGGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 117
Db      1 YGGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 60

QY      118 KKNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 174
Db      61 KKNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 6
US-09-765-271-68
; Sequence 68, Application US/09765271
; Patent No. 6887663
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,271
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/536,784
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-271-68

Query Match      68.6%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.7e-55;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      58 YGGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 117
Db      1 YGGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKEENKPTFDVSK 60

QY      118 KKNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 174
Db      61 KKNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 7
US-09-765-272A-68
; Sequence 68, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
```

```
; APPLICATION NUMBER: US/09/765,272A
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin J. Hymel
; REGISTRATION NUMBER: 45,414
; REFERENCE/DOCKET NUMBER: FB340P2C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 610-5790
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272A-68

Query Match      68.6%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.7e-55;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 YKGELEKGYQDFGWEISGFEKKDAGYVINLSKDTFKVPFKIEBKKEENKPTPDVSK 117
DB 1 YKGELEKGYQDFGWEISGFEKKDAGYVINLSKDTFKVPFKIEBKKEENKPTPDVSK 60

QY 118 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPK 174
DB 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPK 117

RESULT 8
US-09-134-001C-3868
; Sequence 3868, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3868
; LENGTH: 778
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3868

Query Match      13.5%; Score 121; DB 2; Length 778;
Best Local Similarity 26.8%; Pred. No. 0.0012;
Matches 51; Conservative 26; Mismatches 65; Indels 48; Gaps 10;

QY 6 DPAR--NTTVKEFILNKDTEGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK---- 59
DB 601 DFKSVPNVQDREYQKAEDSNAQSLKP-----ITIGNGKQIKQOSVSGTKVLPKSHKVM 656
QY 60 --GELEKGYQDFGWEISGFEKKDAGYVIN--LSKDTFKVPFKIEBKKEENKPTPDVSK 105
DB 657 TDGELTWP-DMTGWTEDVLAFEDLTFKIKVSTKNGFVTVNGSISKGQIILK----- 705
QY 106 EENKPTPDVVS-----KKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNN 161
DB 706 ---NKDKIEVLSAEDTDDQEKTDDESDSKKDKADEHNSSTSKN-----DKS 756
```

```
QY 162 NISSKSTNN 171
DB 757 NADSKNDSDD 766
```

## RESULT 9

```
US-09-710-279-652
; Sequence 652, Application US/09710279
; Patent No. 6703432
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 652
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-652

Query Match      13.3%; Score 119; DB 2; Length 746;
Best Local Similarity 27.0%; Pred. No. 0.0018;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

QY 20 KDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQDFGW-- 71
DB 585 EDSVNAQSLKP----ITIGNGKQIKQOSVSGTKVLPKSHKVMKMTDGLTWP-DMTGWT 639

QY 72 -EISGFE-----GKDDAGYVIN--LSKDTFKVPFKIEBKKEENKPTPDVVS---K 117
DB 640 EDVLAFEDLTFKIKVSTKNGFVTVNGSISKGQIILK-----NKDKIEVLSAED 686

QY 118 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 171
DB 687 TDDQEKTDDESDSKKDKADEHNSSTSKN-----DKSNADSKNDSDD 734
```

## RESULT 10

```
US-09-248-796A-16224
; Sequence 16224, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA
; FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16224
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16224

Query Match      11.6%; Score 104; DB 2; Length 347;
Best Local Similarity 19.9%; Pred. No. 0.022;
Matches 39; Conservative 42; Mismatches 83; Indels 32; Gaps 6;

QY 4 VKDFARN-TTVKEFILNKDTEGEVSELKPHR-----VTVTIQNGKMSSTIVSE 50
```

```
Db 151 LKTELSPPLKQKFLSYDQVRKLEMTREKLVOAWKDCGLIDFDVDQMLKSLPLMIDE 210
Qy 51 EDFILPVYKGELEKGYQPDGWSISGF-----EGKKDAGYVINLSKDTFIKPVFKKIBEK 104
Db 211 IDTIPRIYND--KQWVVVATSSLQVYQTDLSSSEISGWEDDLSENYRTGPFVKTLQDL 268
Qy 105 KREENKPTPDVSKKNDPNVNSQLN-----ESHRKEDLQREHSQKSDSTKQVTVATVL 158
Db 269 REE-----WKASKEQANPKKEBENLNQKPVAKQKQKPNSTKKQKQTKQKQTKKITKPKT 323
Qy 159 DKNNISSKSTNNPNK 174
Db 324 SKRMLEGISTNIIINK 339

RESULT 11
US-09-976-594-726
; Sequence 726, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 726
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Qy 14 KEFILNKDTGVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
Db 584 EKVMVKDKPKVTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623
Qy 74 SGFEGKKDAGYVINLSKDTFIKPVFKKIBEEKKEENKPTFDVSKKNDPNVNSQLNESH 133
Db 624 QATDVKPKAAKEKTVKKTVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKEBKP 675
Qy 134 RKEDLQRE-----EHSQKSDSTKDV 153
Db 676 KKEEVKKEVKKEIKKEBKEKPKKEV 700

RESULT 12
US-09-538-092-1135
; Sequence 1135, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Qy 14 KEFILNKDTGVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
Db 584 EKVMVKDKPKVTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623
Qy 74 SGFEGKKDAGYVINLSKDTFIKPVFKKIBEEKKEENKPTFDVSKKNDPNVNSQLNESH 133
Db 624 QATDVKPKAAKEKTVKKTVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKEBKP 675
Qy 134 RKEDLQRE-----EHSQKSDSTKDV 153
Db 676 KKEEVKKEVKKEIKKEBKEKPKKEV 700

RESULT 13
US-09-949-016-10237
; Sequence 10237, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10237
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10237

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2522;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Qy 14 KEFILNKDTGVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
Db 638 EKVMVKDKPKVTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 677
Qy 74 SGFEGKKDAGYVINLSKDTFIKPVFKKIBEEKKEENKPTFDVSKKNDPNVNSQLNESH 133
Db 678 QATDVKPKAAKEKTVKKTVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKEBKP 729
Qy 134 RKEDLQRE-----EHSQKSDSTKDV 153
Db 730 KKEEVKKEVKKEIKKEBKEKPKKEV 754

RESULT 14
US-09-134-001C-5157
; Sequence 5157, Application US/09134001C
```

```
Db 151 LKTELSPPLKQKFLSYDQVRKLEMTREKLVOAWKDCGLIDFDVDQMLKSLPLMIDE 210
Qy 51 EDFILPVYKGELEKGYQPDGWSISGF-----EGKKDAGYVINLSKDTFIKPVFKKIBEK 104
Db 211 IDTIPRIYND--KQWVVVATSSLQVYQTDLSSSEISGWEDDLSENYRTGPFVKTLQDL 268
Qy 105 KREENKPTPDVSKKNDPNVNSQLN-----ESHRKEDLQREHSQKSDSTKQVTVATVL 158
Db 269 REE-----WKASKEQANPKKEBENLNQKPVAKQKQKPNSTKKQKQTKQKQTKKITKPKT 323
Qy 159 DKNNISSKSTNNPNK 174
Db 324 SKRMLEGISTNIIINK 339

RESULT 11
US-09-976-594-726
; Sequence 726, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 726
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Qy 14 KEFILNKDTGVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
Db 584 EKVMVKDKPKVTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623
Qy 74 SGFEGKKDAGYVINLSKDTFIKPVFKKIBEEKKEENKPTFDVSKKNDPNVNSQLNESH 133
Db 624 QATDVKPKAAKEKTVKKTVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKEBKP 675
Qy 134 RKEDLQRE-----EHSQKSDSTKDV 153
Db 676 KKEEVKKEVKKEIKKEBKEKPKKEV 700

RESULT 12
US-09-538-092-1135
; Sequence 1135, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2468;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Qy 14 KEFILNKDTGVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
Db 584 EKVMVKDKPKVTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 623
Qy 74 SGFEGKKDAGYVINLSKDTFIKPVFKKIBEEKKEENKPTFDVSKKNDPNVNSQLNESH 133
Db 624 QATDVKPKAAKEKTVKKTVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKEBKP 675
Qy 134 RKEDLQRE-----EHSQKSDSTKDV 153
Db 676 KKEEVKKEVKKEIKKEBKEKPKKEV 700

RESULT 13
US-09-949-016-10237
; Sequence 10237, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10237
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10237

Query Match
Best Local Similarity 10.9%; Score 97.5; DB 2; Length 2522;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

Qy 14 KEFILNKDTGVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
Db 638 EKVMVKDKPKVTETKPSVTEKEVPSKEEPS-----PV-KAEVA-----EK 677
Qy 74 SGFEGKKDAGYVINLSKDTFIKPVFKKIBEEKKEENKPTFDVSKKNDPNVNSQLNESH 133
Db 678 QATDVKPKAAKEKTVKKTVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKEBKP 729
Qy 134 RKEDLQRE-----EHSQKSDSTKDV 153
Db 730 KKEEVKKEVKKEIKKEBKEKPKKEV 754

RESULT 14
US-09-134-001C-5157
; Sequence 5157, Application US/09134001C
```



```
Best Local Similarity 21.9%; Pred. No. 0.61;
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;

QY 11 TTVEFILNKOTGR-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-RGY 66
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 336 SAITEFQNVQPTNERKMTLDQTKYVYVESVENNESMDTFVKH-----PIKGTMLNGKKY 390
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 QP-----DGEWISGFEGKK-----D 81
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 MYMETNDYDKDFVMEGQQRVETISKDAKNNTRTIIPPVVEGKTLTYDAIVKVHVKTIDYD 450
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 AGYVNL-SKOTFIKPVFKKIEKKKEENKPTFDV-----SKKDNPOVNHSQLNESHK 135
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 GQYHVIYDVKEAFTKANTQSKNKEQDQNSAKKEATPATPSKPTSPVKEKESQKQDSQKD 510
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 136 EDLQ-----REHSQKSDSKQVT-ATVLDKNNISSKSTNNPNK 174
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 511 DNKQLPSVEKENDASSESKDKTKPATKPTKGEVSSITT--PTK 552
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 18
US-09-248-796A-17646
; Sequence 17646, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 17646
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-17646

Query Match 10.4%; Score 93; DB 2; Length 280;
Best Local Similarity 25.7%; Pred. No. 0.21;
Matches 43; Conservative 26; Mismatches 58; Indels 40; Gaps 10;

QY 18 LNKDTGEVSELPKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGVQPDGWEISGF 77
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 INPIGVHSS-----TYTKT--SVASTFCSKYDFNVFSYASNLSLGF-----ELYSYA 113
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 78 GKGDAGYVNLKOTFIKPVFKKIE-EKKKEENKPTFDVSKKDNPOVNHSQLNESHK 135
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 NKK-----KNSF--PSPHEHHSSEENK---VLKCHPELQRHNLHNLHQR 158
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 136 EDLQREHS-----QKSDTKDVTATVLDKNNISSKSTNNPN 173
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 159 VPIKSHKVEGNRTIINPIQLNDVNYHINPTLLSSNG-STSTTTNNEN 204
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 19
US-08-669-286-7
; Sequence 7, Application US/08669286
; Patent No. 6130060
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, SEIJI
; APPLICANT: SAKURAI, TAKASHI
; APPLICANT: NEZU, JUNI-ICHI
; TITLE OF INVENTION: GENE ENCODING ADSEVERIN
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church

STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/669,286
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/669,286
FILING DATE:

STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/469,253
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/669,286
FILING DATE:

Query Match 10.4%; Score 93; DB 2; Length 715;
Best Local Similarity 21.0%; Pred. No. 0.77;
Matches 37; Conservative 39; Mismatches 58; Indels 42; Gaps 6;

QY 3 VVKDFARNTTVKFIPLNKDTGCVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELE 62
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 IIAIDISNRKMAKLYMVSDAGSM-----RVTVAEENPPSMALLSECFILD--HGAA 290
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 EKGYPDQGEWISGFEGK-----KDAGYVNLKOTFIK-----PVFKK----- 100
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 291 KOIPVWKGKDANPOERKAAMKTAEFLQOMNYSKNTQIQLPEGETFIPKQFFKQFDMRDK 350
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 101 -----IEKKKEENKPTFDVSKKDNPOV--NHSQNLNESHKEDLQREHS 144
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 DQSDGFGKVVYTERKVAQIKOIPFDASKLHSSFPQAAQHNMVDDGSGKVEIRVENN 406
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 20
US-09-469-253-7
; Sequence 7, Application US/09469253
; Patent No. 6184352
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, SEIJI
; APPLICANT: SAKURAI, TAKASHI
; APPLICANT: NEZU, JUNI-ICHI
; TITLE OF INVENTION: GENE ENCODING ADSEVERIN
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church

STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/469,253
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/669,286
FILING DATE:
```

```
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 230-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 715 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-469-253-7

Query Match 10.4%; Score 93; DB 2; Length 715;
Best Local Similarity 21.0%; Pred. No. 0.77;
Matches 37; Conservative 39; Mismatches 58; Indels 42; Gaps 6;

QY 3 VVKDPARNTTVKPEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSSEDFILPVYKGEI 62
Db 239 IADISNRKMAKLYWSDASGM-----RVTVAEENPFPSWMLLSECFILD--HGAA 290
QY 63 EKGYPDGEWISGFEKG-----KDAGYVILSKDTFIK-----PVFKK----- 100
Db 291 KQIFWVGKDANPQERKAAMKTAEBFLQOMYSKNTQIQVLPEGGETPIFKQFFKDWDRK 350
QY 101 -----IEKKKEENKPTFDVSKKONPOV--NHSQLNESHKEDLQREHS 144
Db 351 QDSGDFGKVVYVTEKVAQIKQIPFDASKLHSSPQMAAQHNMYDDGSGKVEIWRVENN 406

RESULT 21
US-09-642-146-7
; Sequence 7, Application US/09642146
; Patent No. 6271353
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, SEIJI
; APPLICANT: SAKURAI, TAKASHI
; TITLE OF INVENTION: NEZU, JUNI-ICHI
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolaach & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/642,146
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/669,286
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 230-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 715 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
US-09-642-146-7

Query Match 10.4%; Score 93; DB 2; Length 715;
Best Local Similarity 21.0%; Pred. No. 0.77;
Matches 37; Conservative 39; Mismatches 58; Indels 42; Gaps 6;

QY 3 VVKDPARNTTVKPEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSSEDFILPVYKGEI 62
Db 239 IADISNRKMAKLYWSDASGM-----RVTVAEENPFPSWMLLSECFILD--HGAA 290
QY 63 EKGYPDGEWISGFEKG-----KDAGYVILSKDTFIK-----PVFKK----- 100
Db 291 KQIFWVGKDANPQERKAAMKTAEBFLQOMYSKNTQIQVLPEGGETPIFKQFFKDWDRK 350
QY 101 -----IEKKKEENKPTFDVSKKONPOV--NHSQLNESHKEDLQREHS 144
Db 351 QDSGDFGKVVYVTEKVAQIKQIPFDASKLHSSPQMAAQHNMYDDGSGKVEIWRVENN 406

RESULT 22
US-08-296-791-5
; Sequence 5, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
US-08-296-791-5
```

```
Query Match 10.3%; Score 92.5; DB 2; Length 1702;
Best Local Similarity 23.9%; Pred. No. 2.9;
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;

QY 2 IVVKDFARNT-----TVKEFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSE---EDFI 54
Db 1206 VVSKNQNTENTTDQPTEREKTAKEVTEKQE--PPQVSAQSPKQSQSETVQPVAVLESN 1263
QY 55 LPVYKGELEKGYQPDGWEISGFEKHDA-GVVINLSKDTFIKPVFKLIEKKKEENKPTF 113
Db 1264 VPTVNAEEVQALQQTQTSATVSTVSTKQAPENSINTGSATAITETAEKDKPKQTAASTE 1323
```

QY 114 DVSKKQNPQVNHSQLNESHKEDLQREHS---QKSDSTKDVATVTLDKNNISSKSTTN 170  
Db 1324 DASQHKANTVADNSVANNSESSEPKRRRSISQPOETSABETTAASTDETTIADNSKRS 1383  
QY 171 NPNK 174  
Db 1384 KPNR 1387

RESULT 23  
US-09-839-996-5  
; Sequence 5, Application US/09839996  
; Patent No. 6642371  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; Falkow, Stanley  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/839,996  
; FILING DATE: 20-Apr-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791  
; FILING DATE: 25-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-09-839-996-5

Query Match 10.3%; Score 92.5; DB 2; Length 1702;  
Best Local Similarity 23.9%; Pred. No. 2.9;  
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;

QY 2 IVVKDPART-----TVKGFILNKDTGEVSELKPHRTVTVIQNGKEMSSIVSE---EDFI 54  
Db 1206 VVSKNTENTTQPTEREKTAKEVTEKTE---PPQVASQASPKQEQSETVQPOAVLESEN 1263

QY 55 LPVYKGELEKGYQFDGWEISGPEGKDA-GYVINLSKDTFIKPVFKKIEKKKEENKPTF 113  
Db 1264 VPTVNNAEVQAOLOQTSAVTSVKQAPENSINTGSAITETAESKDPQTETAASTE 1323

QY 114 DVSKKQNPQVNHSQLNESHKEDLQREHS---QKSDSTKDVATVTLDKNNISSKSTTN 170  
Db 1324 DASQHKANTVADNSVANNSESSEPKRRRSISQPOETSABETTAASTDETTIADNSKRS 1383

QY 171 NPNK 174  
Db 1384 KPNR 1387

RESULT 25  
US-10-645-655-5  
; Sequence 5, Application US/10645655  
; Patent No. 6815182  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; Falkow, Stanley  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/645,655  
; FILING DATE: 20-Aug-2003  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791

RESULT 24  
US-10-080-505-5  
; Sequence 5, Application US/10080505  
; Patent No. 6676948  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme, Joseph W.  
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS  
; FILE REFERENCE: A-59941-1/RPT/DCF/DHR  
; CURRENT APPLICATION NUMBER: US/10/080,505  
; CURRENT FILING DATE: 2002-02-22  
; PRIOR APPLICATION NUMBER: US 08/296,791  
; PRIOR FILING DATE: 1994-10-25  
; PRIOR APPLICATION NUMBER: US 09/839,996  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 5  
; LENGTH: 1702  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae  
US-10-080-505-5

Query Match 10.3%; Score 92.5; DB 2; Length 1702;  
Best Local Similarity 23.9%; Pred. No. 2.9;  
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;

QY 2 IVVKDPART-----TVKGFILNKDTGEVSELKPHRTVTVIQNGKEMSSIVSE---EDFI 54  
Db 1206 VVSKNTENTTQPTEREKTAKEVTEKTE---PPQVASQASPKQEQSETVQPOAVLESEN 1263

QY 55 LPVYKGELEKGYQFDGWEISGPEGKDA-GYVINLSKDTFIKPVFKKIEKKKEENKPTF 113  
Db 1264 VPTVNNAEVQAOLOQTSAVTSVKQAPENSINTGSAITETAESKDPQTETAASTE 1323

QY 114 DVSKKQNPQVNHSQLNESHKEDLQREHS---QKSDSTKDVATVTLDKNNISSKSTTN 170  
Db 1324 DASQHKANTVADNSVANNSESSEPKRRRSISQPOETSABETTAASTDETTIADNSKRS 1383

QY 171 NPNK 174  
Db 1384 KPNR 1387

RESULT 25  
US-10-645-655-5  
; Sequence 5, Application US/10645655  
; Patent No. 6815182  
; GENERAL INFORMATION:  
; APPLICANT: St. Geme III, Joseph W.  
; Falkow, Stanley  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration  
; Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/645,655  
; FILING DATE: 20-Aug-2003  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/296,791

; FILING DATE: 25-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
US-10-645-655-5  
  
Query Match 10.3%; Score 92.5; DB 2; Length 1702;  
Best Local Similarity 23.9%; Pred. No. 2.9;  
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;  
  
QY 2 IVVKDFARNT----TVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSE---EDFI 54  
Db 1206 VSKNQNTENTDQPTEREKTAKVETKQE--PPQVASQASPKQEQSETVQPQAVLESEN 1263  
  
QY 55 LPVYKGELEKGYQFDGWEISGPEGKDA-GYVINLSKDTFIKPVFKKIEKKEENKPTF 113  
Db 1264 VPTVNNAEVQAQLQTQTSATVSTKQAPENSINTGSATATETAEKSDKPKQTETAASTE 1323  
  
QY 114 DVSKKNDPQVNHSQLNESHKEDLQREHS---QKSDSTKQVTVATVLDKNNISSKSTTN 170  
Db 1324 DASQKANTVADNSVANNSESSEPKSRRRRSISQPOETSABEETTAASTDETTIADNSKES 1383  
  
QY 171 NPNK 174  
Db 1384 KPNR 1387  
  
RESULT 26  
PCT-US95-10661A-5  
; Sequence 5, Application PC/TUS9510661A  
; GENERAL INFORMATION:  
; APPLICANT: Washington University, et al.  
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Teet, Albritton & Herbert  
; STREET: 4 Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/10661A  
; FILING DATE: 16-AUG-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/296,791  
; FILING DATE: 25-AUG-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Trecartin, Richard F.  
; REGISTRATION NUMBER: 31,801  
; REFERENCE/DOCKET NUMBER: FP-59941/RFT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1702 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; PCT-US95-10661A-5  
  
Query Match 10.3%; Score 92.5; DB 4; Length 1702;  
Best Local Similarity 23.9%; Pred. No. 2.9;  
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;  
  
QY 2 IVVKDFARNT----TVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSE---EDFI 54  
Db 1206 VSKNQNTENTDQPTEREKTAKVETKQE--PPQVASQASPKQEQSETVQPQAVLESEN 1263  
  
QY 55 LPVYKGELEKGYQFDGWEISGPEGKDA-GYVINLSKDTFIKPVFKKIEKKEENKPTF 113  
Db 1264 VPTVNNAEVQAQLQTQTSATVSTKQAPENSINTGSATATETAEKSDKPKQTETAASTE 1323  
  
QY 114 DVSKKNDPQVNHSQLNESHKEDLQREHS---QKSDSTKQVTVATVLDKNNISSKSTTN 170  
Db 1324 DASQKANTVADNSVANNSESSEPKSRRRRSISQPOETSABEETTAASTDETTIADNSKES 1383  
  
QY 171 NPNK 174  
Db 1384 KPNR 1387  
  
RESULT 27  
US-09-710-279-1888  
; Sequence 1888, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PUS4800S  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1888  
; LENGTH: 299  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-1888  
  
Query Match 10.3%; Score 92; DB 2; Length 299;  
Best Local Similarity 24.4%; Pred. No. 0.29;  
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;  
  
QY 19 NKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFI---LPVYKGELEKGYQFDGWEI 73  
Db 33 NKDT-EKSDKKTTHRIISLIPSNTEILYRLGIGEDIVGVSTVDYDPKVKKGGKQFDAMWL 91  
  
QY 74 SGFE-----GKQDAGYVINLSKDTFIKPV-----FKKIEKK 105  
Db 92 NKEELIKAKPDLILAHESQKNSAGKVLKSLDKGVVYVVKDAQSIDETDTFKSIGOUT 151  
  
QY 106 EENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSQ 145  
Db 152 DREKQAKELVDETKHNVEKIINSVPKHHKKQBFMEVSSK 191  
  
RESULT 28  
US-09-134-001C-5667  
; Sequence 5667, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:



```

; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5667
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5667

Query Match          10.3%; Score 92; DB 2; Length 309;
Best Local Similarity 24.4%; Pred. No. 0.31;
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;

QY 19 NKDTGRVSELKPHRVTVTIONGKMSSTIVSEEDFI-----LPVYKGLEKG-YQFDGWEI 73
    ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 43 NKDT-ERSDKKYHRIISLIPNTIELYRLGIEDIVGVSTVDDYPKDVKKGKKQFDANL 101
    ||||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 74 SGFE-----GKKDAGVINLSKDTFIKPV-----FKKIBKK 105
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 102 NKEELIKAKPDILIAHESQKNSACKVLKSLKDGKVVYVKDAQSIDETVTFKSIQLT 161
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 106 EENKPTFDVSKKONQVNHSQLNESHKEDLQREHSQ 145
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
DB 162 DREKQAKELVDETGHNYEKILINSVPKHHKQKEVPMFVSSK 201
    : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

```

RESULT 29  
US-08-360-673-2  
; Sequence 2, Application US/08360673  
; Patent No. 5679544  
; GENERAL INFORMATION:  
; APPLICANT: Fleer, Reinhard  
; APPLICANT: Fournier, Alain  
; APPLICANT: Yeh, Patrice  
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR  
; TITLE OF INVENTION: PREPARATION AND USE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSES: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd. 3c43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19002  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/360,673  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/FR93/00623  
; FILING DATE: 23-JUN-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 92/07785  
; FILING DATE: 25-JUN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Smith, Julie K.  
; REGISTRATION NUMBER: 38,619  
; REFERENCE/DOCKET NUMBER: ST92040-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (610)454-3839

```

; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-360-673--2

Query Match 10.3%; Score 92; DB 1; Length 561;
Best Local Similarity 23.5%; Pred. No. 0.7;
Matches 43; Conservative 32; Mismatches 72; Indels 3

Qy 4 VKDFARNTTVKFI LNKDTGEVSELKPHRVVTVTIQNGKEMSS-----T
Db 90 VNDIANKIPNRYIIVFKDASADEVAFHQELVSVEHAKALGSLADHDPPTT
Qy 56 PVTYKGELEKGYQDQWEISGFEFGKQDAGVYINLSKDT--PIK--PVFKKIE
Db 150 GVKAHSLLEGIGQ-DSFDIAG----SLSGYVGYFTKEVIDFIRASPLVFEVFE
Qy 106 -EEENKPTPDVSKCKNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVTA
Db 205 FNTQNSAPWGLARISHREKLNLSGPN-----KYLVDYDDAGKGVTA
Qy 164 SSK 166
Db 254 NHK 256

RESULT 30
US-09-640-305-2
; Sequence 2, Application US/09640305
; Patent No. RE37447
; GENERAL INFORMATION:
; APPLICANT: Fleer, Reinhard
; Fournier, Alain
; Yeh, Patrice
; TITLE OF INVENTION: MODIFIED KLUYVEROMYCES YEASTS, THEIR
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd. 3C43
; CITY: Collegeville
; STATE: PA
; COUNTRY: USA
; ZIP: 19002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/640,305
; FILING DATE: 16-Aug-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,673
; FILING DATE: 06-FEB-1995
; APPLICATION NUMBER: WO PCT/FR93/00623
; FILING DATE: 23-JUN-1993
; APPLICATION NUMBER: FR 92/07785
; FILING DATE: 25-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Julie K.
; REGISTRATION NUMBER: 38,619
; REFERENCE/DOCKET NUMBER: ST92040-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610)454-3839
; TELEFAX: (610)454-3808
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids

```

```
;
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-640-305-2

Query Match      10.3%; Score 92; DB 5; Length 561;
Best Local Similarity 23.5%; Pred. No. 0.7; Indels 36; Gaps 8;
Matches 43; Conservative 32; Mismatches 72; Indels 36; Gaps 8;

QY 4 VKDFARTTTVEKILNKGTEVSELSKPHRVTVTIQNGKEMSS-----TIVSEEDFIL 55
Db 90 VNDIANKIPNRYIIVFKDASDEVKFKQELVSEHAKALSLADHPPTATSGHSEF 149
QY 56 PYYKGELEKGYQDGEHSEIGPEGKDGAGVYVNLKDT--PIK--PVFKKIEKK----- 105
Db 150 GVKAHSLGGIQ--DSFDIAG----SLSGYGVGYFTKEVIDFIRRSPLVFEEDSMVFNS 204
QY 106 -EENKPTPDVSKKDNPNVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKN-NI 163
Db 205 FNTQNSAPNGLARISHREKLNLSFN-----KYLDDDGAGGTATVDTGVNV 253
QY 164 SSX 166
Db 254 NHK 256

RESULT 31
US-09-134-001C-3033
; Sequence 3033, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3033
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3033

Query Match      10.1%; Score 91; DB 2; Length 442;
Best Local Similarity 21.7%; Pred. No. 0.64; Indels 82; Gaps 11;
Matches 49; Conservative 30; Mismatches 65; Indels 82; Gaps 11;

QY 20 KDTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 64
Db 90 EDT-EINPQAPEVNTQIEKGDFEATVTVPEVKLDGYKGLIEKQETDLSDEELQ 148
QY 65 -----GYQFDGWEISG--PEGKDGAGVYVNLKSDTKFTK 95
Db 149 ESIDHLSHLAEMVVKEDGAVNGDVTNIDFSG-SVDGEFDDGQAEVYDLEIGSGSFIP 207
QY 96 PVFKKIEEKKKEENK-----PTFDVS-----KKKNPNVNHSQLNE- 131
Db 208 GFEQIEGMKTGDEKDVVTFPEYHAEELAGKEATFTKVNIEKPKDVPFELNDEIANEL 267
QY 132 -----SHRKEDLQREHSQKSDTKDVTATVLDKNNTSSKSTTN 170
Db 268 DSDAENVDEKENLRKELSEKQATEAENT-----EKEEAINKATEN 308

RESULT 32
US-08-910-925-3
; Sequence 3, Application US/08910925
```

```
; Patent No. 6162601
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,925
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0365 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 743 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1684847
; US-08-910-925-3

Query Match      10.1%; Score 90.5; DB 2; Length 743;
Best Local Similarity 23.8%; Pred. No. 1.5;
Matches 38; Conservative 34; Mismatches 51; Indels 37; Gaps 7;

QY 27 ELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGB-----LEKGYQFDGWEI--SGF 76
Db 281 EARPR-----QSMKEKHQVVRNEEHKAEQEGKVAQREBELVETGHNNDVEIEEAGE 335
QY 77 EGKDGAGVYVNLKSDTKFTKPVFKKIEEKKBEENKPTFDVSKKDNPNVNHSQLNESHKRB 136
Db 336 EEEKIGIVHSDAE-----KEQEEEEKQKQMEVQKEEE-----TEVRESEKQ 378
QY 137 DLQREHS---QKSDTKDVTAT--TVLDKNNTSSKSTTN 171
Db 379 DSQPEEVDVLMVENVKHVIADQEVMETNTRVESVEPSEN 418

RESULT 33
US-09-949-016-6261
; Sequence 6261, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

```

; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6261
; LENGTH: 743
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6261

Query Match
Best Local Similarity 10.1%; Score 90.5; DB 2; Length 743;
Matches 38; Conservative 34; Mismatches 51; Indels 37; Gaps 7;

QY 27 ELKPHRVTTTQNGKMSSTIVSEEDFILPVYKGE-----LEKGYQPDGWEI--SGF 76
Db 281 EAPRPR-----QSMKEKEHQVVRNEEHKABQEGKVAQREBELVETGNOHNDVEIEEAGE 335
QY 77 EGKDKAGYVNLKSKDTFIKPVFKKIEBKKEENKPTFDVSKKDNQVNHVSOLNESHKXE 136
Db 336 EEEKEIGIVHSDAE-----KEQEEBKQKQEMKVMWEE-----TEVRESEKQQ 378
QY 137 DLOREHRS---QKSDSTKDQVTA--TVLDKNNISSKSTTNN 171
Db 379 DSQPEVMDVLEMVNKHVIADQEVMTNRVSEVPSEN 418

RESULT 34
US-09-949-016-8288
; Sequence 8288, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8288
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8288

Query Match
Best Local Similarity 10.1%; Score 90.5; DB 2; Length 758;
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 23 GEVSELKPHRVTTTQNGKMSSTIVSE--EDFILPVYKGELEKGYQP-----DGWE--- 72
Db 399 GBELQSLGSQL-----HGKSDSPNVYTEKKETAILERLUTELERKLTPEQQRSLDWERLY 453
QY 73 -----ISGPEGKKDAG-----YVNLKSDTFIKPVFKKIEBKKEENKPTFDVSKKQD 120
Db 454 VBAKQNGKQGTGDKKGGKGRSHRAKNSKETPLGSGV-----KETFDAMKNS 501
QY 121 NPQVNH-----SQLNESHKEDLQREHHSQKSD--STKDVATATVLDK 160
Db 502 KEFVRHHEKIKQAQEA--VKENLKKFSDSVKSTFRHFKDTTKNIIDE 547
```

```

RESULT 35
US-09-710-279-264
; Sequence 264, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 264
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-264

Query Match
Best Local Similarity 10.0%; Score 90; DB 2; Length 785;
Matches 42; Conservative 33; Mismatches 78; Indels 24; Gaps 8;

QY 7 PARNTTVKBFILNKDGTGVSELKPHRVTTTQNGKMS-STIVSEEDFILPVYKGELEKG 65
Db 615 YKRTFVSEYRLSNRGKGKIK----TATITERGNHIVITTTVGEDLMVVTNAGVI---- 667
QY 66 YQFDGWEISGPEGKKDAGY-VINLSKDTFIKPVFKKIEBKKEEN-----KPTPDV 115
Db 668 IRLDVHDIISQ-NGRAAQGVRLMKLGQGFVSTVAKVNEEDDNEENADEAQSTTTTADV 726
QY 116 SKKKNPQVNHVSOLNESHKEDLQREH-HSQKSDSTKDQVATVLDKNNISSKSTTNN 171
Db 727 EEVVD-----DOTPGNAIHTEGDAEMESVEFPENDRRIDIRQDFMDRVNEDIESASDN 779

RESULT 36
US-09-248-796A-21451
; Sequence 21451, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBIC
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 21451
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-21451

Query Match
Best Local Similarity 10.0%; Score 89.5; DB 2; Length 262;
Matches 43; Conservative 26; Mismatches 43; Indels 53; Gaps 9;

QY 20 KDTGVSEL-KPHRVTVT-----IQNGKMSSTIVSEEDF----- 53
Db 52 KSTPKTSPLRKPKPKPTVTPVRKMAKRPSPVTNTPEIKPKBSSESPISSEDFLEMD 111
QY 54 -----ILPVYKGELEKGYQFDGWEISGPEGKKDAGVVLNLSKDTFIKPVFKKIEBKKEEN 109
Db 112 KLTEVPIIE-----FNFDNYD--NDEDKKEE--VVKNSKNE--NQNTIKGMELKPKPKP 160
```

QY 110 KPTFDVSKKDPQVNHSQLNESHKEDLQREHSQKSDSTKDV 154  
Db 161 KPEETSKTKVPQ-----PRK---QKKQKPLSEETV 193

## RESULT 37

US-09-464-483-4  
; Sequence 4, Application US/09464483  
; Patent No. 6228617  
; GENERAL INFORMATION:  
; APPLICANT: Lawlor, Elizabeth J.  
; TITLE OF INVENTION: No. 6228617el tig  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/464,483  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/999,339  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Falk, Stephen T  
; REGISTRATION NUMBER: 36,795  
; REFERENCE/DOCKET NUMBER: GM10085  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2488  
; TELEFAX: 215-994-2222  
; TELEX:

INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 402 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-464-483-4

Query Match 10.0%; Score 89.5; DB 2; Length 402;  
Best Local Similarity 22.1%; Pred. No. 0.79;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 21 DTGEVSELKPHRYVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 64  
Db 53 DETDIKPAQPEVSVTQIEKGDFIFEAIVTVEPVKLGDKYGLKLEIKQETLSDDLEQ 112  
QY 65 -----GYQFDGWEISG--FEGKKDAGYVINLSKDTFIKP 96  
Db 113 AIDHSLGHLAEMVVKEDGVVNGDTVNIIDFSG-SVDGEEFEGQAEGYDLEIGSGSFI-P 170  
QY 97 VFK-----KIEEKE-----EE--NKPTFDVS-----KKONPQVNHSQLNE- 131  
Db 171 GFEEQLGKMKVDEKDVVTFPEYHAEELAGKEATFKTKVNEIKFKEVPELTDEIANEL 230  
QY 132 -----SHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 170  
Db 231 DAEANTVDEYKENLRKLAEQKATDAENV-----EKKEAITKATDN 271

## RESULT 38

US-09-414-664-4  
; Sequence 4, Application US/09414664

; Patent No. 6242249  
; GENERAL INFORMATION:  
; APPLICANT: Lawlor, Elizabeth J.  
; TITLE OF INVENTION: No. 6242249el tig  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/414,664  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/999,339  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Falk, Stephen T  
; REGISTRATION NUMBER: 36,795  
; REFERENCE/DOCKET NUMBER: GM10085  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2488  
; TELEFAX: 215-994-2222  
; TELEX:

INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 402 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-414-664-4

Query Match 10.0%; Score 89.5; DB 2; Length 402;  
Best Local Similarity 22.1%; Pred. No. 0.79;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 21 DTGEVSELKPHRYVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 64  
Db 53 DETDIKPAQPEVSVTQIEKGDFIFEAIVTVEPVKLGDKYGLKLEIKQETLSDDLEQ 112  
QY 65 -----GYQFDGWEISG--FEGKKDAGYVINLSKDTFIKP 96  
Db 113 AIDHSLGHLAEMVVKEDGVVNGDTVNIIDFSG-SVDGEEFEGQAEGYDLEIGSGSFI-P 170  
QY 97 VFK-----KIEEKE-----EE--NKPTFDVS-----KKONPQVNHSQLNE- 131  
Db 171 GFEEQLGKMKVDEKDVVTFPEYHAEELAGKEATFKTKVNEIKFKEVPELTDEIANEL 230  
QY 132 -----SHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 170  
Db 231 DAEANTVDEYKENLRKLAEQKATDAENV-----EKKEAITKATDN 271

## RESULT 39

US-09-198-452A-509  
; Sequence 509, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; TITLE OF INVENTION: and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24

; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 509  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-509

Query Match 10.0%; Score 89.5; DB 2; Length 511;  
Best Local Similarity 24.5%; Pred. No. 1.1;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 94 IKPVFKKIEKKKEENKPTD-----VSKKKNPQVNHSQLNESHRK 135  
DB 95 VKGVFKKTPQARPEVSSPLPSHVQGRPLGLEGRDRIQKRSNPADLGMKRSYSYD 154  
QY 136 EDLQREHSQSDSTKDVTATVLDKNNISSKSTT 169  
DB 155 GDLDRVGHDSNEDSTEDSR---EGGEPSSKSSS 185

RESULT 40

US-09-438-185A-475  
; Sequence 475, Application US/09438185A  
; Patent No. 6822071  
; GENERAL INFORMATION:  
; APPLICANT: Stephens, Richard  
; APPLICANT: Mitchell, Wayne  
; APPLICANT: Kalman, Sue  
; APPLICANT: Davis, Ronald  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Chlamydia pneumoniae Genome Sequence  
; FILE REFERENCE: 018941-000411US  
; CURRENT APPLICATION NUMBER: US/09/438,185A  
; CURRENT FILING DATE: 2002-03-13  
; PRIOR APPLICATION NUMBER: US 60/108,279  
; PRIOR FILING DATE: 1998-11-12  
; PRIOR APPLICATION NUMBER: US 60/128,606  
; PRIOR FILING DATE: 1999-04-08  
; NUMBER OF SEQ ID NOS: 1074  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 475  
; LENGTH: 511  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
; FEATURE:  
; OTHER INFORMATION: CPn0473  
US-09-438-185A-475

Query Match 10.0%; Score 89.5; DB 2; Length 511;  
Best Local Similarity 24.5%; Pred. No. 1.1;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 94 IKPVFKKIEKKKEENKPTD-----VSKKKNPQVNHSQLNESHRK 135  
DB 95 VKGVFKKTPQARPEVSSPLPSHVQGRPLGLEGRDRIQKRSNPADLGMKRSYSYD 154  
QY 136 EDLQREHSQSDSTKDVTATVLDKNNISSKSTT 169  
DB 155 GDLDRVGHDSNEDSTEDSR---EGGEPSSKSSS 185

Search completed: April 24, 2006, 15:03:41  
Job time : 29.0537 secs

**THIS PAGE BLANK (uspio)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 87.1946 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773  
Perfect score: 897  
Sequence: 1 KIVVKDFARNTTVKVFILNK.....ATVLDKNNISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	773	4	US-10-067-385-8
2	897	100.0	2119	3	US-09-769-744A-28
3	897	100.0	2140	4	US-10-282-122A-73670
4	897	100.0	2140	5	US-10-472-928-1180
5	894	99.7	637	5	US-10-617-320-3169
6	615	68.6	117	3	US-09-765-272-68
7	615	68.6	117	6	US-11-106-649-68
8	121	13.5	778	4	US-10-724-972A-5663
9	111.5	12.4	707	4	US-10-282-122A-52942
10	111.5	12.4	775	4	US-10-282-122A-70721
11	110.5	12.3	647	5	US-10-691-672A-3
12	109	12.2	665	3	US-09-820-843A-107
13	108.5	12.1	973	5	US-10-732-923-18783
14	108	12.0	188	5	US-10-691-672A-7
15	107.5	12.0	470	5	US-10-739-930-6262
16	107.5	12.0	948	5	US-10-732-923-4286
17	105	11.7	3127	5	US-10-732-923-22588
18	104.5	11.6	903	4	US-10-282-122A-52328
19	104	11.5	540	5	US-10-732-923-22820
20	103.5	11.5	169	5	US-10-691-672A-2
21	102	11.4	1419	5	US-10-732-923-4285
22	101.5	11.3	564	6	US-11-097-143-12723
23	101.5	11.3	1434	5	US-10-732-923-4235
24	101	11.3	933	4	US-10-437-963-166606
25	100.5	11.2	1350	5	US-10-732-923-8760
26	100.5	11.2	1373	5	US-10-732-923-16976
27	100	11.1	1875	4	US-10-369-493-22285

28	100	11.1	1875	5	US-10-732-923-3334	Sequence 3334, Ap
29	100	11.1	1875	5	US-10-732-923-3335	Sequence 3335, Ap
30	99.5	11.1	1384	4	US-10-473-576-22	Sequence 22, Appl
31	99.5	11.1	1404	4	US-10-473-576-2	Sequence 2, Appl
32	99.5	11.1	3124	5	US-10-732-923-22709	Sequence 22709, A
33	98.5	11.0	2060	4	US-10-381-598A-2	Sequence 2, Appl
34	98	10.9	441	5	US-10-510-812-14	Sequence 14, Appl
35	98	10.9	891	4	US-10-437-963-166609	Sequence 166609, A
36	98	10.9	1419	5	US-10-732-923-4288	Sequence 4288, Ap
37	97.5	10.9	1184	4	US-10-282-122A-53254	Sequence 53254, A
38	97.5	10.9	2468	4	US-10-755-889-615	Sequence 615, App
39	97.5	10.9	2468	5	US-10-489-740-216	Sequence 216, App
40	97.5	10.9	2519	5	US-10-450-763-46995	Sequence 46995, A
41	96.5	10.8	898	4	US-10-425-115-205148	Sequence 205148, A
42	96	10.7	893	4	US-10-282-122A-70831	Sequence 70831, A
43	96	10.7	902	4	US-10-724-972A-6781	Sequence 6781, Ap
44	96	10.7	1939	5	US-10-732-923-3340	Sequence 3340, Ap
45	94.5	10.5	234	3	US-09-764-870-367	Sequence 367, App

ALIGNMENTS

RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Chol, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%;	Score 897;	DB 4;	Length 773;
Best Local Similarity	100.0%;	Pred. No. 2.7e-67;	Mismatches 0;	Indels 0; Gaps 0;
Matches 174;	Conservative	0;	Mismatches	0;
QY	1	KIVVKDFARNTTVKVFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKG	60	
DB	600	KIVVKDFARNTTVKVFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKG	659	
QY	61	ELEKGYQDFGWEISFGEGKKDAGYVINSKDTFFIKVPFKCKEKKERENKPTFDVSKKKD	120	
DB	660	ELEKGYQDFGWEISFGEGKKDAGYVINSKDTFFIKVPFKCKEKKERENKPTFDVSKKKD	719	
QY	121	NPQVNHSQLNESHREKDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK	174	
DB	720	NPQVNHSQLNESHREKDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK	773	

RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Phillip M  
; TITLE OF INVENTION: Proteins

```
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-09-769-744A-28

Query Match      100.0%; Score 897; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 9.5e-67;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKPFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKG 60
Db 1912 KIVVKDFARNTTVKPFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKG 1971
QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKD 120
Db 1972 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKD 2031
QY 121 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
Db 2032 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2085

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-10-282-122A-73670

Query Match      100.0%; Score 897; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 9.6e-67;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKPFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKG 60
Db 1933 KIVVKDFARNTTVKPFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKG 1992
QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKD 120
Db 1993 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKD 2052
QY 121 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
Db 2053 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.8+01)
; US-10-472-928-1180

Query Match      100.0%; Score 897; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 9.6e-67;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKPFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKG 60
Db 1933 KIVVKDFARNTTVKPFILNKDTGEVSELKPHRVTVTTIQNGKEMSSITVSEEDFILPVYKG 1992
QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKD 120
Db 1993 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKVPFKKIEKKKEENKPTFDVSKKD 2052
QY 121 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
Db 2053 NPQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```



```
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-10-617-320-3169

Query Match 99.7%; Score 894; DB 5; Length 637;
Best Local Similarity 99.4%; Pred. No. 3.9e-67; Mismatches 0; Indels 0; Gaps 0;
Matches 173; Conservative 1;

QY 1 KIVKDFARNTTVKGFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKG 60
DB 430 KIVKDFARNTTVKGFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKG 489
QY 61 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSKKD 120
DB 490 ELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSKKD 549
QY 121 NPOVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
DB 550 NPOVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 603

RESULT 6
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

```
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match 68.6%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 2e-44;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSK 117
DB 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSK 60
QY 118 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 174
DB 61 KQNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117

RESULT 7
US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

```
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      68.6%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 2e-44;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 YGELKGYQFGWEISGEGKDGAGYVNLKDTFIKVPFKIIEKKKEENKPTFDVSK 117
Db 1 YGELKGYQFGWEISGEGKDGAGYVNLKDTFIKVPFKIIEKKKEENKPTFDVSK 60
;

QY 118 KDNPNVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNISKSTTNNPNK 174
Db 61 KDNPNVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNISKSTTNNPNK 117
;

RESULT 8
US-10-724-972A-5663
; Sequence 5663, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; PRIOR FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 5663
; LENGTH: 778
; TYPE: PRT
; ORGANISM: S. epidermidis
US-10-724-972A-5663

Query Match      13.5%; Score 121; DB 4; Length 778;
Best Local Similarity 26.8%; Pred. No. 0.15;
Matches 51; Conservative 26; Mismatches 65; Indels 48; Gaps 10;

QY 6 DPAP--NTTVKGFILNKDTGEVSELKPHRVTVTQNGKMSSTIVSEEDFILPVVK--- 59
Db 601 DFKVPNVGDEVQKKAEDSVNAQSLKP-----ITIGNGKQIKQOSVKSQKGVLPKPHSKVLM 656
;

QY 60 --GELEKGYQFGW-----EISGPE-----GKKDAGYVIN--LSKDTFTKVPFKIERKK 105
Db 657 TDGELTMP--DMTGWTKEDVLAFEDLTKIKVSTKNGFVTVNQISKQIILK----- 705
;

QY 106 EENKPTFDVS-----KKDNPNVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKN 161
Db 706 ---NKKIEVLSAEDTDDQKEDSDSNKSKKQKADKEDHSNTSSSTKN-----DKS 756
;

QY 162 NISSKSTTN 171
Db 757 NADSKNDSDD 766
;

RESULT 9
US-10-282-122A-52942
; Sequence 52942, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
```

```
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52942
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Clostridium difficile
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(18)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (29)..(29)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (37)..(37)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (43)..(43)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (54)..(54)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (84)..(84)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (86)..(86)
; OTHER INFORMATION: X=any amino acid
```

FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (359)..(359)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (395)..(385)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (388)..(388)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (396)..(396)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (400)..(400)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (402)..(402)  
OTHER INFORMATION: X=any amino acid  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (404)..(404)  
OTHER INFORMATION: X=any amino acid  
US-10-282-122A-52942

Query Match 12.4%; Score 111.5; DB 4; Length 707;  
Best Local Similarity 26.6%; Pred. No. 0.83;  
Matches 47; Conservative 28; Mismatches 63; Indels 39; Gaps 7;  
QY 2 IVKDTGEVSELKPHRVTTIQNGKEMSTIVSEEDFILPVYK 59  
Db 484 ISIEDDAER-GVKEEDSNQDIDGVVEDK-----TTDKEDYS---NKSDIIEPENK 532  
QY 60 GLEKGYQDQWEISGFEGKQAGYVNLKDTFIKPVFKIEEKKKEEN--KPTFDVSK 117  
Db 533 KSKKAKLFG-----FIKKNEEVEQEEENLNDISPDILDK 569  
QY 118 KDNPNVNSQLNESHKEDLQREHSQKSDTKQVTVATVLDKNNISSKSTNNPNK 174  
Db 570 PVENNQVKEIEQNLKE-IKQEPHQIEBERSVKIEKPINNILDEKVVSSNNESK 625

RESULT 10  
US-10-282-122A-70721  
Sequence 70721, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 70721  
LENGTH: 775  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-10-282-122A-70721

Query Match 12.4%; Score 111.5; DB 4; Length 775;  
Best Local Similarity 24.9%; Pred. No. 0.93;  
Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;  
QY 20 KDTGEVSELKPHRVTTIQNGKEMSTIVSEEDFILPVYK-----GELEKGYQDQW-- 71  
Db 614 EDSVNAQSLKP-----ITINGKQIKQSVKSTKVLPHSKVLMMDTDELTMP-DMTGWTYK 668  
QY 72 -RISGFE-----GKKDAGYVIN--LSKDTFIKPVFK-----KIEKKKEENKPTF 113  
Db 669 EDVLAPEDLTKIKVSTKNGFVTVNSISKQIINKDKLEVSLSAEDTDDQKTDDESS 728  
QY 114 DVSKKKDNPQNVHSQLNESHKEDLQREHSQKSDTKQVTVATVLDKNN 162  
Db 729 DKSKKDKVDENSNASSSSKNEKSNADSKNDSDDSTNETSGS--ERNN 775

RESULT 11  
US-10-691-672A-3  
Sequence 3, Application US/10691672A  
Publication No. US20050112133A1  
GENERAL INFORMATION:  
APPLICANT: DRUILHE, PIERRE  
TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND  
FILE REFERENCE: 02356.0085  
CURRENT APPLICATION NUMBER: US/10/691,672A  
CURRENT FILING DATE: 2003-10-24  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 3.3  
SEQ ID NO 3  
LENGTH: 647  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)..(647)  
OTHER INFORMATION: GLURP MSP3 fusion protein  
US-10-691-672A-3

Query Match 12.3%; Score 110.5; DB 5; Length 647;  
Best Local Similarity 22.8%; Pred. No. 0.91;  
Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;  
QY 14 KEFILNKDTGVESELKPHRVTTIQNGKEM-----SSTIVSEEDFILPVYKGELE 63

```
Db 418 EBAVSEKNAHETVE---HEETVQESNPEKADNGNVSONNENELNEFV-----ESE 468
Qy 64 K-----GQFD-GWEISGF--EGKQDAG-----VYINLSKDTFKVPVKKIE 102
Db 469 KSEHARSKAKASSYDITLGMFGGVPEHKKENMLSHLYVSSKKENISKENDVDLD 528
Qy 103 EKKEENKPTFDVSKKKDQPNVHSQLN-----ESHKREDLQREHSHKSDS 149
Db 529 E-KEEAETEEBELKEKEETESEISEDEBEDEBEDEBEDEBEDEBEDEBEDE 587
Qy 150 TKDVTATVLDKNNISSKSTNN 171
Db 588 KKOMEA-----QNLISKQNNN 604

RESULT 12
US-09-820-843A-107
; Sequence 107, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|3845248
US-09-820-843A-107

Query Match 12.2%; Score 109; DB 3; Length 665;
Best Local Similarity 26.1%; Pred. No. 1.3;
Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;

Qy 5 KDFARNTTVKEFTLNKDTGVSSELKPHRTV-VTIQNGKEMSSITVSEEDFILPVYKGELE 63
Db 114 KDNNNNGTKLEEKNKINKSDL--HRQNELMLQSGK-----NEQDI-----NKNE 158
Qy 64 KGQFDGWRI--SGPFGKQDAGYVINLSKDTFKVPVKKIEBKE-----EENKP 111
Db 159 KGKQ-----DISNSAENKQD-----VKEGVKLEEKKEEKISDDHKVEENCK 202
Qy 112 TPD-----VSKKDNQPNVHSQLNESHKEDLQREHSHKSDS TKDVTATVLDKNNISSK 166
Db 203 SDDHKVEENKSDDHKVEENKSDDHKIEBVKVKEHEBDEE-----DKKEKSE 253
Qy 167 STTNPNK 174
Db 254 NKNKDNK 261

RESULT 13
US-10-732-923-18783
; Sequence 18783, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149

; SEQ ID NO 18783
; LENGTH: 973
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(973)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-18783

Query Match 12.1%; Score 108.5; DB 5; Length 973;
Best Local Similarity 21.7%; Pred. No. 2.2;
Matches 38; Conservative 44; Mismatches 80; Indels 13; Gaps 5;

Qy 5 KDFARNTTVKEF-IILNKDTGB-----VSELKPHRTVVTIQNGKEMSSITVSEEDFILPV 57
Db 568 KDIRDNKSSSFTLVNNNTNQNNGNSYDDKLETYININNDQKGECNSTYKT---LIQH 623
Qy 58 YKGELEKGYQFDGWEISGFEGKQDAGYVINLSKDTFKVPVKKIEBKEENKPTFDVSK 117
Db 624 RSGSKESGKTHYTKINNISRREHARGYVSLSKDNNIEVRNDNIEVRNDNIEV 683
Qy 118 KKNQPNVHSQLNESHKEDL-QREHSHKSDS TKDVTATVLDKNNISSKSTNN 171
Db 684 RNDNIEMRNDISNDQNEKNISETNSFNKSEYT-FVTATSNKSKDNNINKSSND 737

RESULT 14
US-10-691-672A-7
; Sequence 7, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 7
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(188)
; OTHER INFORMATION: MSP3a to MSP3f
US-10-691-672A-7

Query Match 12.0%; Score 108; DB 5; Length 188;
Best Local Similarity 23.2%; Pred. No. 0.32;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

Qy 37 IQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKQDAG-----YVINLS 89
Db 15 VLKAKEASS-----YDYIL-----GHEFGGVPEHKKENMLSHLYVSSKD 55
Qy 90 KDTFKVPVKKIEBKEE-----ENKPTFDVSKKKNQPNVHSQLNESHKKE 136
Db 56 KENISKENDVDLDEKEEAEETEHEEKEEETESEISEDEBEDEBEDEBEDEBEDEBNK 115
Qy 137 DLQREHSHKSDS TKDVTATVLDKNNISSKSTNN 171
Db 116 EQEKSQSNENNDQKKDMEA-----QNLISKQNNN 145

RESULT 15
US-10-739-930-6262
; Sequence 6262, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
```



```
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52328
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52328

Query Match          11.6%; Score 104.5; DB 4; Length 903;
Best Local Similarity 21.9%; Pred. No. 4.4;
Matches 44; Conservative 40; Mismatches 62; Indels 55; Gaps 9;

QY 8 ARNTTVKPFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQ 67
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 494 SKNEKSKQ-VLNRNTQLKNEKSKQLPKINVELKEEKNKQIIKEHN-----ELEKG-- 542

QY 68 PDGWEISGEGKADAGYVNLKDTFIKPVFKIEKKEENKPTFDVSV-----KKQDN 121
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 543 -----GKQPKQKVNIELNKEK-VKQVFNESIESKNEKSKQVVRNTQLKNEKSKQL 592

QY 122 PQVN-----HSQLNESHKKE-----DLORE-----BHSQ-KSDSTKDV 153
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 593 PKINVELKEEKSKQISKEHNEELKEKKKPKINVELKEEKSKQIIKEHNEKNEKSKQT 652

QY 154 TATVLDKNNISSKSTTNNPNK 174
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 653 PKVNVLEKKEKAGHVFNESIK 673

RESULT 19
US-10-732-923-22820
; Sequence 22820, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22820
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-732-923-22820

Query Match          11.6%; Score 104; DB 5; Length 540;
Best Local Similarity 25.3%; Pred. No. 2.6;
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 11 TTVKPFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFG 70
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 27 TTVKA-VVETKVEEESKP-----EGVEKSASFKEESDFADLKESEKK----- 70

QY 71 WEISGEGKADAGYVNLKDTFIKPVFKK---IEEKKEENKPTFDVSKKQNPQVNHIS 127
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 71 -ALSDLSKLEEAIVDN-----TLTKTKKESSPMKGEVVKPEAEVKKKE--BAABE 123

QY 128 QLNESHKEDLOREHSQKSDSTKDVTVTLVDKNNISS 165
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
```

```
Db 124 KVEBEKKSEAVVTEEPKAETVEAVVTEIIPKEEVT 161

RESULT 20
US-10-691-672A-2
; Sequence 2, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691.672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 2
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(169)
; OTHER INFORMATION: MSP3 amino acids 212-380
US-10-691-672A-2

Query Match          11.5%; Score 103.5; DB 5; Length 169;
Best Local Similarity 25.2%; Pred. No. 0.67;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 41 KEMSSTIVSEEDFILPVYKGELEKGYQDGEWEISGF--EGKKDAG-----YVINLSKDTF 93
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   |||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 1 KEASS-----YDYL-----GWFGGVPHEKKENMLSHLYVSSKKENI 41

QY 94 IKPVFKITEEKEEENKPTFDVSKKQNPQVNHSQLN-----BSHRKEDLQR 140
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 42 SKENDDVLDL-KEBEAEETEEELKNEEETESEISEDEEEEBEKEEKEKKEQEK 100

QY 141 EEHSQKSDSTKDVTVTLVDKNNISSKSTTNN 171
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 101 EQSNNENDQKKDMEA-----QNLISKNQNNN 126

RESULT 21
US-10-732-923-4285
; Sequence 4285, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 4285
; LENGTH: 1419
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-4285

Query Match          11.4%; Score 102; DB 5; Length 1419;
Best Local Similarity 22.0%; Pred. No. 13;
Matches 52; Conservative 37; Mismatches 77; Indels 70; Gaps 11;

QY 4 VKDFARNTTVKPFILNKDTGEVSEL-----KPHRVTVTIQNGKMSSTIVSEEDFI 54
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 1019 VFDSSNMGFFNYIINLGGYPPEELEYDNKAGDKIHYCTLSLQIKDKIQKFLNEETFL 1078

QY 55 LPVY-----KGELEKGYQDGEWEISGEGKKD-----AGYVINLSKDTFIKPVFKK 100
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 1079 KTKGYGSFEKISLAINNSIDHY-----FSHMKDNLRLVCEPGRYVMAASSTLAVKIIGKR 1134
```

[illegible]

RESULT 22

US-11-097-143-12723

; Sequence 12723, Application US/11097143

; Publication No. US20050208558A1

; GENERAL INFORMATION:

; APPLICANT: Venter, J. Craig

; APPLICANT: et al.

; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID

; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE

; TITLE OF INVENTION: DROSOPHILA GENES.

Query Match	11.3%	Score 101.5	DB 6	Length 564
Best Local Similarity	24.5%	Pred. No. 4.4		
Matches	34	Conservative 29	Mismatches 51	Indels 25
Gaps	5			
Qy	51	EDFTLPVYKGELEKGYQDGN-----RISGPEGKQAGYVI-----NLSKQTFIK	95	
Db	78	EDLTPUSEGRFSK--VPDGVWDHRDEHDGHVQEPGSGEALDDHDEHDHDDHEDDEE	135	
Qy	96	PVFKIKBIKKKEENKPT-----FDVSKKONPQWNHSQLNESHRKEDLOREHHSQKSDS	149	
Db	136	PLTSELESELESESEPTFEDEPADEEYVEDEDEENNA--GENITAEADASEEEEEENDD	193	
Qy	150	TKQVTATVLDKNNTSSKST	168	
Db	194	EGTVEATVATATATAT	212	

RESULT 23  
US-10-732-923-4235  
; Sequence 4235, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732, 923

```

; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 4235
; LENGTH: 1434
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-4235

```

## RESULT 24

```

US/10-437-963-166606
; Sequence 166606, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kowalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Mol
; FILE NO OF INVENTION: Plants and Uses Ther
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-03-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 166606
; LENGTH: 933
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530
US/10-437-963-166606

```

	Query Match	11.3%	Score 101,	DB 4,	Length 933,
	Best Local Similarity	23.5%	Pred. No. 9.1,		
	Matches 43,	Conservative 39,	Mismatches 75,	Indels 26,	Gaps 7;
Qy	3	VVKDFARNTVTKSEFILNK-----DTGEVSELKPHRVTVTIQNGKEMSSTVSEED	52		
Db	212	LVSNFSCRDVLVEBFVLLKTRPLSGWDITLGEIPECPSSLPPFV-----VSSDLVIPED	266		
Qy	53	FILPVYK--GELEKGYQFDGWEISGPGKQDAGYVNLNSKDTIKPVPFKKI--EEKKEE	108		
Db	267	MVPVPVKFLGPTTKDEHLKFLTLQ--NGRQRNVPASLGSDDIPVR-VYPEIVFLPKKKKS	323		
Qy	109	NKPTFDVSKKQKQPNVHSGQNLNSHRKEDLQREHSQKGDSTKQVTATVLDKNNISKST	168		
Db	324	VKP-----SSDDDDDDVEVDVDEI GENAEAE EEEBEGADEENDDSDSDSSSSSSNS	379		

```
QY 169 TNN 171
Db 380 SDS 382

RESULT 25
US-10-732-923-8760
; Sequence 8760, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8760
; LENGTH: 1350
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) - (1350)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-8760

Query Match 11.2%; Score 100.5; DB 5; Length 1350;
Best Local Similarity 23.0%; Pred. No. 16;
Matches 46; Conservative 37; Mismatches 74; Indels 43; Gaps 10;

QY 5 KQPARNTTVKBFILNKDTGEVSELKPHRVTVTQNGK-----EWSSTIVSEDFILPVY 58
Db 846 KKLKNSSTTKKFLKNN-----KKHMFPLDLGERKSKWKVMYNTACTKTNKKKAILYGW 897

QY 59 KGELEKGYOPDQW-----EISGFEGKQDAGYVNLNSKDTFKPVFKLEEKKEE---N 109
Db 898 KAERAGGHDPQFPNNEKLELEKIEKWN-NYHINQOK---IKBIIESQTEKEDFEKIIN 953

QY 110 KPTFDVSKKKD-----NPQV-NHSQLNESHKEDLORE-----EHSQKSDSTKQVT 154
Db 954 IHEFLTHAKNIYLNILNPNLNDNSIGDSIEEKQVSKDIDEDGNEGKYSGEFNGI- 1012

QY 155 ATVLDKNNISSKSTNNPNK 174
Db 1013 VNIYKSNVSGSSETRNDSK 1032

RESULT 26
US-10-732-923-16976
; Sequence 16976, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 16976
; LENGTH: 1373
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-16976

Query Match 11.2%; Score 100.5; DB 5; Length 1373;
Best Local Similarity 21.7%; Pred. No. 16;
Matches 40; Conservative 33; Mismatches 54; Indels 57; Gaps 9;

QY 16 FILNKDTGEVSE---LKPFRVTVTQNGKEMSSSTIVSEDFILPVYKGELEKGYOPDQW 71
Db 418 VYNEDRGVWEESIIMLPFHVQILJ-----LSATV-----PNY-----LEFADW 457

QY 72 -----EISGFEGKK-----DAGYVNLNSKDTFKPVFK-----KIEBKKEE 108
Db 458 VGFTKQKEVISISTKCRPVPLLLHYTVYDSVVLVMDKKNKFYSSAFKEIYVYKIREKQEAN 517

QY 109 NKPIFDVSKKDNQVNHSQLNESH-RKEDLOREHSQKSDSTKQVTATVLDKNNISSK 167
Db 518 NKNTKQITSGSNTTSSNLKNNYVDSKNKYLTITNNKENDNT-----QNNINNNN 568

QY 168 TNN 171
Db 569 NNNN 572

RESULT 27
US-10-369-493-22285
; Sequence 22285, Application US/10369493
; Publication No. US2003023367SA1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22285
; LENGTH: 1875
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22285

Query Match 11.1%; Score 100; DB 4; Length 1875;
Best Local Similarity 24.4%; Pred. No. 26;
Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;

QY 4 VKDFARNTTVKBFILNKDTGEVSELKPHRVTVTQNGKEMSSSTIVSEDF-----ILPVY 58
Db 647 ISQITRESTENMSLINK---BIQDIYDSKSDISIKLGKESRILAEERFKLLSNTLDLT 703

QY 59 KGE---LEKGYOPDQWIEISGFEGKKDA-----GYVNLNSKDTFKPVFKKIEK--- 104
Db 704 KAENDQLRRFDY-----LQNTILKQDSTKTHETLNEYVSKSKLSIVETELMLKBEQKUR 759

QY 105 -----KEEENKPTFDVSKKDNFQVNHSQLNESH-RKEDL---QRBEHSQKSDSTKQV 153
Db 760 VHLEKNLQELNK-----LSPEKDSLRIMVTQLOTQKEREDELEETRSCKCKKIDLEDA 815

QY 154 TA-----TVLDKNNISSKSTNNPN 173
Db 816 LSELKKETSQKDHIIKQLEEDNNSN 840

RESULT 28
US-10-732-923-3334
; Sequence 3334, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
```



; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 3334  
; LENGTH: 1875  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-732-923-3334

Query Match 11.1%; Score 100; DB 5; Length 1875;  
Best Local Similarity 24.4%; Pred. No. 26;  
Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;  
QY 4 VKDFARNTTVKEFILNKDTGEVSELKPHRVTVTQNGKMSSTIVSEEDP-----ILPVY 58  
Db 647 ISQITRESTENSLNKK---EIQDLYDSKDISIKLGKESRIILABERFKLLSLTDLT 703  
QY 59 KGE---LEKGYQFDGWEISGFGKKDA-----GVVNLKSDTTPKVPFKKIEBK----- 104  
Db 704 KAENDQLRRFDY----LQNTILKQDSKTHETLNEYVSCSKLSIVETELLNLKEQKLR 759  
QY 105 -----KEENKPTFDVSKKONQOVNHSQLNESHK-KEDL---OREHHSQKSDTKDV 153  
Db 760 VHLKLNKQELNK---LSPKDSLRIMVTQLTKQKREDLLEETRSQCKKIDLEDA 815  
QY 154 TA-----TVLDKNNISSKSTTNPN 173  
Db 816 LSELKKTBSQKDHHLKQLEEDNNSN 840

RESULT 29  
US-10-732-923-3335  
; Sequence 3335, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15 (52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 3335  
; LENGTH: 1875  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-732-923-3335

Query Match 11.1%; Score 100; DB 5; Length 1875;  
Best Local Similarity 24.4%; Pred. No. 26;  
Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;  
QY 4 VKDFARNTTVKEFILNKDTGEVSELKPHRVTVTQNGKMSSTIVSEEDP-----ILPVY 58  
Db 647 ISQITRESTENSLNKK---EIQDLYDSKDISIKLGKESRIILABERFKLLSLTDLT 703  
QY 59 KGE---LEKGYQFDGWEISGFGKKDA-----GVVNLKSDTTPKVPFKKIEBK----- 104  
Db 704 KAENDQLRRFDY----LQNTILKQDSKTHETLNEYVSCSKLSIVETELLNLKEQKLR 759  
QY 105 -----KEENKPTFDVSKKONQOVNHSQLNESHK-KEDL---OREHHSQKSDTKDV 153  
Db 760 VHLKLNKQELNK---LSPKDSLRIMVTQLTKQKREDLLEETRSQCKKIDLEDA 815  
QY 154 TA-----TVLDKNNISSKSTTNPN 173  
Db 816 LSELKKTBSQKDHHLKQLEEDNNSN 840

RESULT 30  
US-10-473-576-22  
; Sequence 22, Application US/10473576  
; Publication No. US20040101884A1

; GENERAL INFORMATION:  
; APPLICANT: INCYTE CORPORATION  
; APPLICANT: LU, DYUNG AINA M.  
; APPLICANT: ARVIZU, CHANDRA S.  
; APPLICANT: GANDHI, AMEENA R.  
; APPLICANT: HAPALIA, APRIL J.A.  
; APPLICANT: DING, LI  
; APPLICANT: LU, YAN  
; APPLICANT: RAMKUMAR, JAYALAXMI  
; APPLICANT: SWARNAKAR, ANITA  
; APPLICANT: TANG, Y. TOM  
; APPLICANT: YUE, HENRY  
; APPLICANT: TRAN, BAO  
; APPLICANT: LEE, SOO YUEN  
; APPLICANT: WARREN, BRIDGET A.  
; APPLICANT: NGUYEN, DANNIEL B.  
; APPLICANT: THANGAVELU, KAVITHA  
; APPLICANT: YAO, MONIQUE G.  
; APPLICANT: ELLIOTT, VICKI S.  
; APPLICANT: BAUGHN, MARIAH R.  
; APPLICANT: EMERLING, BROOKE M.  
; APPLICANT: LAL, PREETI G.  
; APPLICANT: GIETZEN, KIMBERLY J.  
; APPLICANT: BECHA, SHANYA D.  
; APPLICANT: MARQUIS, JOSEPH P.  
; APPLICANT: KABLE, AMY B.  
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT  
; FILE REFERENCE: PF-0921 USN  
; CURRENT APPLICATION NUMBER: US/10/473,576  
; CURRENT FILING DATE: 2003-09-29  
; PRIOR APPLICATION NUMBER: PCT/US02/09809  
; PRIOR FILING DATE: 2002-03-29  
; PRIOR APPLICATION NUMBER: US 60/280,387  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: US 60/282,335  
; PRIOR FILING DATE: 2001-04-05  
; PRIOR APPLICATION NUMBER: US 60/286,663  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: US 60/285,484  
; PRIOR FILING DATE: 2001-04-19  
; PRIOR APPLICATION NUMBER: US 60/350,702  
; PRIOR FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: US 60/351,749  
; PRIOR FILING DATE: 2002-01-25  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PERL Program  
; SEQ ID NO 22  
; LENGTH: 1384  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID NO: 7506096CD1  
US-10-473-576-22

Query Match 11.1%; Score 99.5; DB 4; Length 1384;  
Best Local Similarity 21.7%; Pred. No. 20;  
Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;  
QY 14 KEFILNKDTGEVSELKPHRVTVTQNGKMSSTIVSEEDPILPVYKGELEKGYQF----- 68  
Db 1013 KEFILNQEORISQIK-KBIERTQORMKEMESVMKQEQYIATQYKEAIDLGQELRLTRE 1071  
QY 69 -----DGWEISGFGKKDA-----GVVNLK----- 90  
Db 1072 QVQNSHTLAEARHQVQAQREIERLSSELEDMQLSKEDKAGNHNLAELGASKVREAH 1131  
QY 91 -----DTFKPVPFKKIEBK-----BENKPTFDVSKKONQOVNHSQLNESHK 136  
Db 1132 LEARMQABIKKLSAEVESLKEAYHWMISHQENHAKWKIS--ADSKSVQQLNQLEKA 1189  
QY 137 DLQREHSQKSDTKDVTATVLDKNNI 163

	Query Match	11.1%;	Score 99.5;	DB 4;	Length 1404;
	Best Local Similarity	21.7%;	Pred. No. 20;		
	Matches	45;	Conservative 32;	Mismatches 67;	Indels 63; Gaps 7;
Qy	14	KETFLNKDTGEVSELKPHRVTVTTQNCKEMSSSTTVSEEDFILPVYKGELSKGYOF-----	68		
		: : :     :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :			
Dd	1033	KEPTMLQNEQEISLQ-KETERTQRMKESVMKQEQVIATQYKAIDLGQELRLTRE	1091		
Qy	69	-----DGWIEISGFEGKKDA-----GYVINLSK-----	90		
Dd	1092	OVO NSHT ELAEARHCOOVAOAEIERLSSELEDMKOLSKEDAHGNHLIAELGASKVKREAH	1151		
		: :			

Matches 45; Conservative 25; Mismatches 67; Indels 39; Gaps 9;

QY 1 KIVVKDPAFNTTVKFEILNKDTGEVSELKPHRVTVTI---QNGKMSSTIVSEEDFILPV 57  
Db 1887 KVTEDK-GITTVYVERIQSVNDAPRVDIDELKTIIVYDTCRE-----IVPS 1934

QY 58 YKGELEKGYQFDG--WEISFEGKKDAGYVNLNLSKDTFIKPVFKKIEE-----KKEBE 108  
Db 1935 RKQQLPP-BOFIGQDMQYTGHK-----TEKDGIITYYIKKVENAVPAKQLKKTGH 1983

QY 109 NKPTFDVSKKKDNPQVNHSQLNESHKE--DLQREHSHQSDSKDVTATVLDKNNI 163  
Db 1984 N--TQSESQPKHTPOVKQQLVKNVKEQBSIEKSEHTDMHVSLEPGETGTANKGL 2038

RESULT 34  
US-10-510-812-14  
; Sequence 14, Application US/10510812  
; Publication No. US20050176097A1  
; GENERAL INFORMATION:  
; APPLICANT: Rasmussen, Michael Dolberg  
; APPLICANT: Jorgensen, Steen Troels  
; APPLICANT: Andersen, Jens Tonne  
; APPLICANT: Olsen, Peter Bjarke  
; TITLE OF INVENTION: Improved Bacillus host cell  
; FILE REFERENCE: 10296.204-US  
; CURRENT APPLICATION NUMBER: US/10/510,812  
; CURRENT FILING DATE: 2004-10-07  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 14  
; LENGTH: 441  
; TYPE: PRT  
; ORGANISM: Bacillus licheniformis  
US-10-510-812-14

Query Match 10.9%; Score 98; DB 5; Length 441;  
Best Local Similarity 31.6%; Pred. No. 6.5;  
Matches 43; Conservative 18; Mismatches 35; Indels 40; Gaps 9;

QY 6 DPA-RNTTVKFEILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEK 64  
Db 302 DYAFENFTKKEFY-----AKGAQVKGNE-TVTVDKGGKETEGLVITKAFSMPYKNGE-BK 354

QY 65 GYQFDGWEISFEGKKDAGYVNLNLSKDTFIKPVFKKIEE-----KKEENKPTF---D 114  
Db 355 NYKAK-----VTLNKDELTAAPV-KKGEKVGILTASYKGEERKDYGLGSD 397

QY 115 VS-----KKKDNQPN 125  
Db 398 VSGVNLVTKEDDEKAN 413

RESULT 35  
US-10-437-963-166609  
; Sequence 166609, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 166609

; LENGTH: 891  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_65300C.1.pap  
US-10-437-963-166609

Query Match 10.9%; Score 98; DB 4; Length 891;  
Best Local Similarity 22.8%; Pred. No. 15;  
Matches 42; Conservative 39; Mismatches 75; Indels 28; Gaps 7;

QY 3 VKDPAFNTTVKFEILNK-----DTGVSSELKPHRVTVTIQNGKMSSTIVSEED 52  
Db 199 LVSNFNSCRDLVEFVLLKIRPLSSGWDITLGIPEDCPSLPPFV-----VSSDLVIPED 253

QY 53 FILPVVK--GELEKGYQFDGWEISFEGKKDAGYVNLNLSKDTFIK---PVPKKIREKKEE 107  
Db 254 MVRPVKFLGPTTKDEHLKFLTQ--NGKQNVFASLGSDIPRVHPFIVPALKKKKSV 311

QY 108 ENKPTFDVSKKKDNPQVNHSQLNESHREKDLQREHSHQSDSKDVTATVLDKNNISSKS 167  
Db 312 --KPSSD---DDDDVLEEDVDEIGENAEEREGEGADEENDSSDDSDSSNSSSDN 365

QY 168 TTNN 171  
Db 366 SSDS 369

RESULT 36  
US-10-732-923-4288  
; Sequence 4288, Application US/10732923  
; Publication No. US20050108791A1  
; GENERAL INFORMATION:  
; APPLICANT: Edgerton, Michael D  
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES  
; FILE REFERENCE: 38-15(52796)C  
; CURRENT APPLICATION NUMBER: US/10/732,923  
; CURRENT FILING DATE: 2003-12-10  
; PRIOR APPLICATION NUMBER: 10/310,154  
; PRIOR FILING DATE: 2002-12-04  
; NUMBER OF SEQ ID NOS: 24149  
; SEQ ID NO 4288  
; LENGTH: 1419  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-10-732-923-4288

Query Match 10.9%; Score 98; DB 5; Length 1419;  
Best Local Similarity 20.0%; Pred. No. 27;  
Matches 50; Conservative 34; Mismatches 68; Indels 98; Gaps 10;

QY 4 VKDPAFNTTVKFEILNKDTGEVSEL-----KPHRVTVTIQNGKMSSTIVSEEDPI 54  
Db 1019 VFDMSNNMGFNYIINLGGYPPELEYDNKAKHDKHYCTLSLQBIKKDIQKFLNEETPL 1078

QY 55 LPVYKGELEKGYQFDGWEISFEGKKDAGYVNLNLSKDTFI-----94  
Db 1079 KTKY-----GYSEKISLAINNSIDHYFSHMKONLAVICEPGSNMVA 1122

QY 95 -----KPVFKKI-----EKKKEENKPTFDVSKKKDNQPN 125  
Db 1123 SSTLAVKIIGKRRTFFQIGIMLKDLKHYDPLNFAPQENKKQDETAKINEN-NDNNNDNN 1181

QY 126 HSQLESKRE-----DL-----QREHSHQSDSKDVTATVLDK--NNIS 164  
Db 1182 DNNINNNNNKGGQGNIMNDLIITSTNDSTSKNDHS--SSQVIQNVCTTRDKEGDNIK 1240

QY 165 -SKSTNNPN 173  
Db 1241 INTHTINPN 1250

RESULT 37



```
; APPLICANT: Hysseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 46995
; LENGTH: 2519
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1040)..(1091)
; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,
; OTHER INFORMATION: accession number BL00412D, p-value=2.432e-11, raw score of 16.54
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1919)..(2122)
; OTHER INFORMATION: Neuraxin and MAP1B proteins domain identified by PFam,
; OTHER INFORMATION: accession name MAP1B_neuraxin, E-value=1.9e-59, PFam score of 190
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2519)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-46995
```

```
Query Match          10.9%; Score 97.5; DB 5; Length 2519;
Best Local Similarity 24.8%; Pred. No. 61;
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 14 KEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQPDGWEI 73
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 635 EKVMVKDKPKVKTETKPSVTEKEVPKKEPS-----PV-KAEVA-----EK 674

QY 74 SGFEGKKDAGYVINLSKDTFIKVPFKIEKKKEENKPTFDVSKKQDNPNQVNHSQLNESH 133
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 675 QATDVKPKAAKEKTVKGETKVP-----EDKKEKEKPKKEVAKCEDKTPI---KKBEKP 726

QY 134 RKEDLQRE-----EHSQKSDSTKDV 153
   :| :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 727 KKEEVKKEVKKEIKKEEKKPKKEV 751
```

Search completed: April 24, 2006, 15:43:38  
Job time : 89.1946 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 13.6242 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897

Sequence: 1 KLVKDFARNTTVKFIILNK.....ATVLDKNNISKSTNNPNK 174

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New.\*  
1: /SIDSS/ptodata/2/pubppaa/US08\_NEW\_PUB.pep.\*  
2: /SIDSS/ptodata/2/pubppaa/US06\_NEW\_PUB.pep.\*  
3: /SIDSS/ptodata/2/pubppaa/US07\_NEW\_PUB.pep.\*  
4: /SIDSS/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep.\*  
5: /SIDSS/ptodata/2/pubppaa/US09\_NEW\_PUB.pep.\*  
6: /SIDSS/ptodata/2/pubppaa/US10\_NEW\_PUB.pep.\*  
7: /SIDSS/ptodata/2/pubppaa/US11\_NEW\_PUB.pep.\*  
8: /SIDSS/ptodata/2/pubppaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119	13.3	746	US-10-793-626-652	Sequence 652, App
2	112.5	12.5	354	US-11-189-817-2	Sequence 2, Appli
3	110.5	12.3	651	US-11-128-660-1	Sequence 1, Appli
4	95	10.6	886	US-11-087-099-11456	Sequence 11456, A
5	93.5	10.4	501	US-10-485-517-381	Sequence 381, App
6	93.5	10.4	645	US-10-485-517-244	Sequence 244, App
7	93	10.4	493	US-11-096-568A-3070	Sequence 3070, Ap
8	93	10.4	493	US-11-096-568A-3071	Sequence 3071, Ap
9	93	10.4	510	US-11-096-568A-3069	Sequence 3069, Ap
10	92	10.3	299	US-10-793-626-1888	Sequence 1888, Ap
11	90	10.0	785	US-10-793-626-254	Sequence 264, App
12	89.5	10.0	1036	US-11-096-568A-28315	Sequence 28315, A
13	89.5	10.0	1070	US-11-096-568A-28314	Sequence 28314, A
14	89.5	10.0	1276	US-11-096-568A-28313	Sequence 28313, A
15	89.5	10.0	1694	US-11-052-554A-83	Sequence 83, Appl
16	87	9.7	140	US-11-096-568A-4771	Sequence 4771, Ap
17	86	9.6	443	US-11-188-298-1015	Sequence 1015, Ap
18	86	9.6	700	US-11-188-475-74	Sequence 74, Appl
19	85.5	9.5	734	US-11-072-512-2272	Sequence 2272, Ap
20	85	9.5	394	US-10-793-626-552	Sequence 552, App
21	85	9.5	472	US-10-793-626-658	Sequence 658, App
22	85	9.5	636	US-10-485-517-170	Sequence 170, App
23	85	9.5	700	US-11-196-475-66	Sequence 66, Appl
24	85	9.5	708	US-11-196-475-76	Sequence 76, Appl
25	85	9.5	720	US-10-793-626-2058	Sequence 2058, Ap

## ALIGNMENTS

## RESULT 1

US-10-793-626-652  
; Sequence 652, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PUS480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 652  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-652

Query Match	13.3%	Score 119;	DB 6;	Length 746;
Best Local Similarity	27.0%	Pred. No. 0.044;		
Matches	47;	Conservative	24;	Mismatches 57;
				Indels 46;
				Gaps 9;

  

QY	20	KDTGEVSELKPHRVTVTTIQNGKEMSSSTIVSEDFILPVYK-----GELEKGYQFDGW---71
DB	585	EDSVNAOSLKP-----ITINGKQIKQOSVSGTKVLPHSKVLMTMDGELTWP-DMTGWTK 639

  

QY	72	-EISGFE-----GKKDAGYVIN--LSKDTFIKVPFKIBKKEENKPTFDVS---K 117
DB	640	EDVLAFEDLTCLKVSTKNGFVTVQSIKGIK-----NKDKLEVSLSAED 686

  

QY	118	KDNQPNVHNSQLNSHRKEDLQREHSQKSTKDVTVATVLDKNNISKSTNN 171
DB	687	TDDQDEKTDSDSNKSKKDKADEHSNTSSSTKN-----DKSNADSKNDSDD 734

## RESULT 2

US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US20060030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRUILHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMIDIUM FALCIPARUM





QY 67 QP-----DQWEISGEGKK-----D 81  
Db 238 MVMETNDYWDKDFWVGQRTISKAKNTRTIIPYVEGKTLVDALVVKVTIDYD 297  
QY 82 AGVINL-SKDTFIKPVFKKIEKKEENKPTFDV-----SKKONPQVNHSQLNESHK 135  
Db 298 GQYHVRIVDKAFTKANTDKSNKGEQDNSAKKEATPATPSKPTSPVEKESQKQDSQKD 357  
QY 136 EDLQ-----REHSQKSDTKQVT-ATVLDKNNISSKSTNNPNK 174  
Db 358 DNKQLPSVEKENDASSEGKDKTPATKPTKGEVSSSTT--PTK 399

RESULT 6

US-10-485-517-244  
; Sequence 244, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, Simon  
; APPLICANT: Mond, James  
; FILE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629W0  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; CURRENT FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118825.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 244  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-244

Query Match 10.4%; Score 93.5; DB 6; Length 645;  
Best Local Similarity 21.9%; Pred. No. 4.5;  
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;  
QY 11 TTVKGFILNKDTGE-VSELKPHRVTV--TIQNGKMSSTIVSEEDFILPVYKGELE-KGY 66  
Db 327 SAITBFQVQPTNEKMTLDQTKYVYVESVENNESMOTFVKH-----PIKTGMLGKKY 381  
QY 67 QP-----DQWEISGEGKK-----D 81  
Db 382 MVMETNDYWDKDFWVGQRTISKAKNTRTIIPYVEGKTLVDALVVKVTIDYD 441  
QY 82 AGVINL-SKDTFIKPVFKKIEKKEENKPTFDV-----SKKONPQVNHSQLNESHK 135  
Db 442 GQYHVRIVDKAFTKANTDKSNKGEQDNSAKKEATPATPSKPTSPVEKESQKQDSQKD 501  
QY 136 EDLQ-----REHSQKSDTKQVT-ATVLDKNNISSKSTNNPNK 174  
Db 502 DNKQLPSVEKENDASSEGKDKTPATKPTKGEVSSSTT--PTK 543

RESULT 7

US-11-096-568A-3070  
; Sequence 3070, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptides  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 3070  
; LENGTH: 493

; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(493)  
; OTHER INFORMATION: Ceres Seq. ID no. 15172486  
US-11-096-568A-3070

Query Match 10.4%; Score 93; DB 7; Length 493;  
Best Local Similarity 26.8%; Pred. No. 3.5;  
Matches 57; Conservative 24; Mismatches 78; Indels 54; Gaps 11;

QY 3 VVKDPARNTTVKPFIL-----NKDTGE-----VSELKPHR---VVTIIONGKEM--SSTI 47  
Db 34 VVSDPSVKTTKKKIIRKVPKKVVGGEASKSLVSEPKDENQGDSTQSSGKQTADANTI 93  
QY 48 VSEE-----DFILPVYKGELEKGYQFDGWEISGFEKGKADAGYVI----- 86  
Db 94 VTEKKPGKVPVKKIKITPVSK---KKDETADSNKTETLSDKKDEGNVAVQAQDDTQST 150  
QY 87 ---NLSKDTFIKPVFKKIEE--KKEEENKPTFDVSKKDNFQVNHSQLNESHKEDLQRE 141  
Db 151 GKQTANADTTVPEVKTKGVVPPKKQSKTPT---SEKRDN--TADSSKTETKSKDDKKE 205  
QY 142 EHSQKSDTKQVTATVLDKNNISSKSTNNPNK 174  
Db 206 ER-----VTGEKSGAKTDKLGKASDKDQVTVNVKKG 233

RESULT 8

US-11-096-568A-3071  
; Sequence 3071, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptide  
; FILE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 3071  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Glycine max  
; NAME/KEY: misc feature  
; LOCATION: (1)-(493)  
; OTHER INFORMATION: Ceres Seq. ID no. 16625552  
US-11-096-568A-3071

Query Match 10.4%; Score 93; DB 7; Length 493;  
Best Local Similarity 26.8%; Pred. No. 3.5;  
Matches 57; Conservative 24; Mismatches 78; Indels 54; Gaps 11;

QY 3 VVKDPARNTTVKPFIL-----NKDTGE-----VSELKPHR---VVTIIONGKEM--SSTI 47  
Db 34 VVSDPSVKTTKKKIIRKVPKKVVGGEASKSLVSEPKDENQGDSTQSSGKQTADANTI 93  
QY 48 VSEE-----DFILPVYKGELEKGYQFDGWEISGFEKGKADAGYVI----- 86  
Db 94 VTEKKPGKVPVKKIKITPVSK---KKDETADSNKTETLSDKKDEGNVAVQAQDDTQST 150  
QY 87 ---NLSKDTFIKPVFKKIEE--KKEEENKPTFDVSKKDNFQVNHSQLNESHKEDLQRE 141  
Db 151 GKQTANADTTVPEVKTKGVVPPKKQSKTPT---SEKRDN--TADSSKTETKSKDDKKE 205  
QY 142 EHSQKSDTKQVTATVLDKNNISSKSTNNPNK 174  
Db 206 ER-----VTGEKSGAKTDKLGKASDKDQVTVNVKKG 233

RESULT 9



	Matches	43;	Conservative	34;	Mismatches	76;	Indels	25;	Gaps	7;
Qy	3	VVKDPARNTTVKFEILNK----	DTGEVS-----	ELKPHRVTTVTIQNGKEMESTIVS	BEDF	53				
		: : : : : : : : : : :	:	:	:	:	:	:	:	:
Db	484	VPKESTGCTSSNKKIIVKVAETG	DTSIDPSAKANQT	PAKTIVKKKI	KVAKRKA	VEIDN	543			
		: : : : : : : : : : :	:	:	:	:	:	:	:	:
Qy	54	ILPVYKLELKGYPDGVWEISGFEG	KK--	DAGVINLSKDTPIKPVFVKIK	BEKKEERNKP	111				
		: : : : : : : : : : :	:	:	:	:	:	:	:	:
Db	544	KM----	DGDSKKDGSDSEKKVMEV-	GKKS	SDSGSV-----	EMKPTAESLEDVDKENASK	592			
		: : : : : : : : : : :	:	:	:	:	:	:	:	:
Qy	112	TFDYSKKKONQVNHSQLNESHRED	LQREH----	SQKSDTKDV	TATVLDKN	NISSK	166			
		: : : : : : : : : : :	:	:	:	:	:	:	:	:
Db	593	TVDDYKQETGS	PDTTCKKCGASSSSK	DTGTGD	KKAEKKNSETMS	EHGKKI	DRNNITDEK	650		
		: : : : : : : : : : :	:	:	:	:	:	:	:	:

```

RESULT 13
US-11-096-568A-28314
; Sequence 28314, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: therapy
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28314
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1070)
; OTHER INFORMATION: Ceres Seq. ID no. 2712009
US-11-096-568A-28314

```

	Query Match	10.0%;	Score 89.5;	DB 7;	Length 1070;
	Best Local Similarity	24.2%;	Pred. No.18;		
	Matches	43;	Conservative 34;	Mismatches 76;	Indels 25; Gaps 7;
Qy	3	VVKDFARNTVTKFILANK--DTGEVS-----ELKPHRVTVTTQNGKMSSTIVSRDP	53		
Db	518	VPKBESTGTSGNKIKLVKVAETGSDTSPSAKANEGTTPAKTIVKKIKRIVAKRVKVAIDN	577		
Qy	54	ILPVYKGELEKGYQFDGWEISGFGKK--DAGYYVINLSKDTFKVPKFKCKEKEEENKP	111		
Db	578	KM-----DGSKDGSDSKKWEV--GKSSDSGSV-----EMPTAESLEDVDEKASK	626		
Qy	112	TFVSVKKKNPQVNHSQLNESHKREDLQREH-----SQKSDSTKDVTATVLDKNNISSK	166		
Db	627	TVDVYQETGSPDPTKKCGGASSSSKDPTGTGBDKKAEEKNNSETMSGKKIDRNNNTDK	684		

```

RESULT 14
US-11-096-568A-28313
; Sequence 28313, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TYPE OF INVENTION: Theaby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28313
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1276)

```

## RESULT 18

```
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2272
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2272

Query Match
Best Local Similarity 9.5%; Score 85.5; DB 7; Length 734;
Matches 39; Conservative 42; Mismatches 58; Indels 41; Gaps 11;

QY 20 KDTG-EVSELKPHRVTV-----TTONKEMSSITVSEE--DFILPVVK 59
Db 168 EDTGIYVSVDPNFNSIAAKDGRIRGDRILQINGEDVQNRREAVALLSNDECKRIVLLVAR 227
QY 60 GLEKGYQDGEISGFKGKAGYVNLKDTF-----IKPVFKTEE-KCEEENKP 111
Db 228 PEIQLD--EGW---LEDERNE-FLEELNLEMEHEHNEAQPANEVEQPKQEEBEG 279
QY 112 TFDVSKKDNPPQVNHSQLNESHR-KEDLQREHSQKSDSTKDTATVL-DKNNISSKSTT 169
Db 280 ITDTATSSN---NHEKDSGVRGRTDSLRNDSSESQENAAEDPNSTSLKSKRDLGQSQDT 336

RESULT 20
US-10-793-626-552
; Sequence 552, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 552
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-552

Query Match
Best Local Similarity 9.5%; Score 85; DB 6; Length 394;
Matches 44; Conservative 13; Mismatches 45; Indels 44; Gaps 10;

QY 10 NTTVKEPILNKDTGEVSELKPHRVTVTTONKEMSSITV---VSEEDPILPVVKGELEKGY 66
Db 55 NINVSOFENPKTHE-SEV--HTASSRANNGRFPVVPFKLASEHIDITQVKG-----105
QY 67 QPDGWEISGEG---KDGAGYVNLKDTFIKPVFKIEBKKEENKPTFDVSKKKNPQ 123
Db 106 -----TGFEGRVTKDQNTQINNPNP-----QBEKEKPKQT-----DKKH-S 142
QY 124 VNHQ-LNESHKEDLQREHSQKSD 148
Db 143 TNHCDPLHQSSTK-----NEHSPLSN 163

RESULT 21
US-10-793-626-658
; Sequence 658, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-658

Query Match
Best Local Similarity 9.5%; Score 85; DB 6; Length 472;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 101 IEKKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQ-----REEHSQKSDS---149
Db 3 MEENKQPNKE--NMSNKDDNA---THLNDSHRNDELLEFRNKNAQRORRRRIDNQSK 56
QY 150 TKQVTAT-----VLDRNNISSKSTTNPNK 174
Db 57 EKDATSTQSQLETKPMDKPLDNHKS--HNQNK 86

RESULT 22
US-10-485-517-170
; Sequence 170, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 170
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-170

Query Match
Best Local Similarity 9.5%; Score 85; DB 6; Length 636;
Matches 35; Conservative 35; Mismatches 61; Indels 28; Gaps 7;

QY 17 ILNKDTGEVSELKPHRVTVTTONKEMSSITVSEEDPILPVVKGELEKGYQPDGWEISGF 76
Db 27 IVTKDYSKQSNAGSKNGTLDLSRYLNSALYYLEDYI--IYAILGTLTKRYE-----GD 78
QY 77 EGKGAGYVNLKDTFIKPVFKK-----IEBKCEE-ENKPTFDVSKKKNPQVNHSQLN- 130
Db 79 NIYKEA-----KDRLEKVLREDOYLLERKKSQYEDYKQYANYKKENFRTDLKMANP 131
QY 131 ESHRKEDLQREHSQKSDSTKDV-----TATVLDRNN 162
Db 132 HKYNLEELSMKEYNELQALKRALDDPHREVKDIKDKNS 170

RESULT 23
US-11-196-475-66
; Sequence 66, Application US/11196475
```

```
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; US-11-196-475-66

Query Match          9.5%; Score 85; DB 7; Length 700;
Best Local Similarity 24.4%; Pred. No. 25;
Matches 44; Conservative 33; Mismatches 53; Indels 50; Gaps 9;

QY 33 VVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVINLSK-- 90
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 258 ITETIENLRDLQEKATDEE-----HKKEIES--QVDAKKQKEELDCKA---INLDKRAQ 306

QY 91 -----DFTIKPVFKIEEKKKEENKP-TFDVSKKKNPQVN-----HS 127
   :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 307 QKLDSAEADNLVQRDITREKIQEDINEINKKNLPKPGDVSSPKVKQIKESLEDLQE 366

QY 128 QLNES---HRKEDLQREHSQSD-----STKQVATATVLDKNNISSKSTTNNPNK 174
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 367 QLKETGDNQKEIEKQIEIKSDEKLKSKDDKASKGKALDLDLR-ELNSKASSKESK 425

RESULT 24
US-11-196-475-76
; Sequence 76, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 76
; TYPE: PRT
```

```
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-76

Query Match          9.5%; Score 85; DB 7; Length 708;
Best Local Similarity 27.6%; Pred. No. 25;
Matches 43; Conservative 22; Mismatches 53; Indels 38; Gaps 8;

QY 33 VVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGVINLSK-- 90
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 258 ITETIENLRDLQEKATDEE-----HKKEIES--QVDAKKQKEELDCKA---INLDKRAQ 306

QY 91 -----DFTIKPVFKIEEKKKEENKP-TFDVSKKKNPQVNHSQLNESHK 135
   :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 307 QKLDSAEADNLVQRDITREKIQEDINEINKKNLPKPGDVSSPKVKQQL---QIKES--L 361

QY 136 EDLQRE-----EHSQKSDSTKDVATVLDKNNISSK 166
   ||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 362 EDLQQLKEAGDENQKEIEKQIEIKRDELLASK 397

RESULT 25
US-10-793-626-2058
; Sequence 2058, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2058

Query Match          9.5%; Score 85; DB 6; Length 720;
Best Local Similarity 28.0%; Pred. No. 26;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 101 IEEKKEENKPTFDVSKKD---NPOVNSQLNESHKEDLQREHHSQKSDSTK---D 152
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 38 LEEEQIKALDKFKKASQAKQTNKQNTQNNHOKSNKQNSNDKEKQKQSKNSKPTKKKEQN 97

QY 153 VTATVLDKNNISSKSTTNNPNK 174
   :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 98 NKGKQKNKNTKNQKNK 119

RESULT 26
US-11-096-568A-26839
; Sequence 26839, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/056,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 26839
; LENGTH: 974
; TYPE: PRT
```

```
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(974)
; OTHER INFORMATION: Ceres Seq. ID no. 13623377
US-11-096-568A-26839

Query Match
Best Local Similarity 20.0%; DB 7; Length 974;
Matches 36; Conservative 33; Mismatches 54; Indels 57; Gaps 8;

QY 2 IVKDPFARNT-----TVKEFILNKDTGVSSELKPHRVTVTTIONGKEMSTIVS 49
Db 791 IVKDFKDVLRIDISPSASLDAIKEWL-----DTTDLKYYESRLNL-NWRPILKTIID 843
QY 50 BEDFILPVYKGELEKGYQDQWEISGFEG-----KKGAGYVINLSKDTFIKPVFK 100
Db 844 DP-----QKFDIDGWEFLNMEASDSETEDETSQGVV-----PSDAE 882
QY 101 IEKKEEENKPTFDVSKKONQVNHSQLNESHK--EDLQRE-----EHSQKSDSTKD 152
Db 883 PESESEDDSDSESLVESDDDESEDESEBEGKGTWEELEREASNADREHGAESDSBEE 942

RESULT 27
US-11-096-568A-26838
; Sequence 26838, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 26838
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1055)
; OTHER INFORMATION: Ceres Seq. ID no. 13623376
US-11-096-568A-26838

Query Match
Best Local Similarity 20.0%; DB 7; Length 1055;
Matches 36; Conservative 33; Mismatches 54; Indels 57; Gaps 8;

QY 2 IVKDPFARNT-----TVKEFILNKDTGVSSELKPHRVTVTTIONGKEMSTIVS 49
Db 872 IVKDFKDVLRIDISPSASLDAIKEWL-----DTTDLKYYESRLNL-NWRPILKTIID 924
QY 50 BEDFILPVYKGELEKGYQDQWEISGFEG-----KKGAGYVINLSKDTFIKPVFK 100
Db 925 DP-----QKFDIDGWEFLNMEASDSETEDETSQGVV-----PSDAE 963
QY 101 IEKKEEENKPTFDVSKKONQVNHSQLNESHK--EDLQRE-----EHSQKSDSTKD 152
Db 964 PESESEDDSDSESLVESDDDESEDESEBEGKGTWEELEREASNADREHGAESDSBEE 1023

RESULT 28
US-11-096-568A-26837
; Sequence 26837, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
```

```
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 26837
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1057)
; OTHER INFORMATION: Ceres Seq. ID no. 13623375
US-11-096-568A-26837

Query Match
Best Local Similarity 20.0%; DB 7; Length 1057;
Matches 36; Conservative 33; Mismatches 54; Indels 57; Gaps 8;

QY 2 IVKDPFARNT-----TVKEFILNKDTGVSSELKPHRVTVTTIONGKEMSTIVS 49
Db 874 IVKDFKDVLRIDISPSASLDAIKEWL-----DTTDLKYYESRLNL-NWRPILKTIID 926
QY 50 BEDFILPVYKGELEKGYQDQWEISGFEG-----KKGAGYVINLSKDTFIKPVFK 100
Db 927 DP-----QKFDIDGWEFLNMEASDSETEDETSQGVV-----PSDAE 965
QY 101 IEKKEEENKPTFDVSKKONQVNHSQLNESHK--EDLQRE-----EHSQKSDSTKD 152
Db 966 PESESEDDSDSESLVESDDDESEDESEBEGKGTWEELEREASNADREHGAESDSBEE 1025

RESULT 29
US-11-191-374-16
; Sequence 16, Application US/11191374
; Publication No. US20050260673A1
; GENERAL INFORMATION:
; APPLICANT: Hresko, Michelle Coutu
; APPLICANT: McLaird, Merry B.
; APPLICANT: Williams, Deryck J.
; APPLICANT: Frevert, Anita M.
; APPLICANT: Chiapelli, Brandi
; APPLICANT: Baublite, Catherine
; APPLICANT: Kloeck, Andrew P.
; APPLICANT: Davila-Aponte, Jennifer A.
; APPLICANT: Bradley, John D.
; APPLICANT: Xu, Siqun
; TITLE OF INVENTION: NEMATODE PAN AND ZP RECEPTOR-LIKE
; FILE REFERENCE: 12557-015001
; CURRENT APPLICATION NUMBER: US/11/191,374
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/771,708
; PRIOR FILING DATE: 2004-02-04
; PRIOR APPLICATION NUMBER: US 60/444,771
; PRIOR FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-191-374-16

Query Match
Best Local Similarity 23.0%; DB 7; Length 1065;
Matches 51; Conservative 36; Mismatches 56; Indels 79; Gaps 13;

QY 6 DPARTTVTKFELNKDTG-----EVSELKPHRVTVTTIONGKEMSTIVSEEE----- 51
Db 570 DFKNLVTKFELNKDTG-----DTGASAAANVPETK-HRVSGTVVVEGDKSKSOLLATKKVKKPTIKN 625
QY 52 -----DFILPV-----YKG-----ELEKGYQDQWEISGFEGKKD----- 81
Db 626 TEHRRAPSTVFIQPPVEVKAEAIQTCINYGKIQVINNGEPFSG--VIFVKNKFTCRV 683
```





```
Query Match          9.4%; Score 84.5; DB 7; Length 1069;
Best Local Similarity 23.0%; Pred. No. 46;
Matches 51; Conservative 36; Mismatches 56; Indels 79; Gaps 13;

QY 6 DPARNTTVKEFILNKDTG-----EVSSELKPHRVTVTIQNGKMSSTIVSE----- 51
Db 574 DFKNLCTVKEI---DTGASAAANVPETK-HRVSGTVVEGKDSKSQLATKKVKKPTIKN 629
QY 52 -----DFILPV-----YKG-----ELEKGYQFDGWEISGPEGKKD----- 81
Db 630 TEHRRAPESVTPIGPPVEVKAEAIQTCINYEIGIKVQINNGEPFSG--VIFVKNKPTDTCRV 687
QY 82 -----AGVINLSKDTFIKPV-FKKIEEKKERENKPTFDVSKKKNPQVNHSQLNESH 133
Db 688 EVANSNAATLVGLPKDFGMRPISLDNIDNETGKNK-----TKKGE-----ETP 732

QY 134 RKEDLQREHSQKSDSTKDTATVLDKNNISKST-----TNN 171
Db 733 LKDEI---EBFRQKQAAEFRCDCGLVDLLNGTYKSTVVIQTNN 772

RESULT 33
US-11-191-375-17
; Sequence 17, Application US/11191375
; Publication No. US20050260674A1
; GENERAL INFORMATION:
; APPLICANT: Hresko, Michelle Coutu
; APPLICANT: McLaird, Merry B.
; APPLICANT: Williams, Deryck J.
; APPLICANT: Frevert, Anita M.
; APPLICANT: Chiapelli, Brandi
; APPLICANT: Baublite, Catherine
; APPLICANT: Kloeck, Andrew P.
; APPLICANT: Davila-Aponte, Jennifer A.
; APPLICANT: Bradley, John D.
; APPLICANT: Xu, Siqun
; TITLE OF INVENTION: NEMATODE PAN AND ZP RECEPTOR-LIKE
; FILE REFERENCE: 12557-015001
; CURRENT APPLICATION NUMBER: US/11/191,375
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/771,708
; PRIOR FILING DATE: 2004-02-04
; PRIOR APPLICATION NUMBER: US 60/444,771
; PRIOR FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1069
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-191-375-17

Query Match          9.4%; Score 84.5; DB 7; Length 1069;
Best Local Similarity 23.0%; Pred. No. 46;
Matches 51; Conservative 36; Mismatches 56; Indels 79; Gaps 13;

QY 6 DPARNTTVKEFILNKDTG-----EVSSELKPHRVTVTIQNGKMSSTIVSE----- 51
Db 574 DFKNLCTVKEI---DTGASAAANVPETK-HRVSGTVVEGKDSKSQLATKKVKKPTIKN 629
QY 52 -----DFILPV-----YKG-----ELEKGYQFDGWEISGPEGKKD----- 81
Db 630 TEHRRAPESVTPIGPPVEVKAEAIQTCINYEIGIKVQINNGEPFSG--VIFVKNKPTDTCRV 687
QY 82 -----AGVINLSKDTFIKPV-FKKIEEKKERENKPTFDVSKKKNPQVNHSQLNESH 133
Db 688 EVANSNAATLVGLPKDFGMRPISLDNIDNETGKNK-----TKKGE-----ETP 732

QY 134 RKEDLQREHSQKSDSTKDTATVLDKNNISKST-----TNN 171
Db 733 LKDEI---EBFRQKQAAEFRCDCGLVDLLNGTYKSTVVIQTNN 772
```

```
Query Match          9.4%; Score 84.5; DB 7; Length 1069;
Best Local Similarity 23.0%; Pred. No. 46;
Matches 51; Conservative 36; Mismatches 56; Indels 79; Gaps 13;

QY 6 DPARNTTVKEFILNKDTG-----EVSSELKPHRVTVTIQNGKMSSTIVSE----- 51
Db 574 DFKNLCTVKEI---DTGASAAANVPETK-HRVSGTVVEGKDSKSQLATKKVKKPTIKN 629
QY 52 -----DFILPV-----YKG-----ELEKGYQFDGWEISGPEGKKD----- 81
Db 630 TEHRRAPESVTPIGPPVEVKAEAIQTCINYEIGIKVQINNGEPFSG--VIFVKNKPTDTCRV 687
QY 82 -----AGVINLSKDTFIKPV-FKKIEEKKERENKPTFDVSKKKNPQVNHSQLNESH 133
Db 688 EVANSNAATLVGLPKDFGMRPISLDNIDNETGKNK-----TKKGE-----ETP 732

QY 134 RKEDLQREHSQKSDSTKDTATVLDKNNISKST-----TNN 171
Db 733 LKDEI---EBFRQKQAAEFRCDCGLVDLLNGTYKSTVVIQTNN 772
```

```
RESULT 34
US-11-191-588-17
; Sequence 17, Application US/11191588
; Publication No. US20050282222A1
; GENERAL INFORMATION:
; APPLICANT: Hresko, Michelle Coutu
; APPLICANT: McLaird, Merry B.
; APPLICANT: Williams, Deryck J.
; APPLICANT: Frevert, Anita M.
; APPLICANT: Chiapelli, Brandi
; APPLICANT: Baublite, Catherine
; APPLICANT: Kloeck, Andrew P.
; APPLICANT: Davila-Aponte, Jennifer A.
; APPLICANT: Bradley, John D.
; APPLICANT: Xu, Siqun
; TITLE OF INVENTION: NEMATODE PAN AND ZP RECEPTOR-LIKE
; FILE REFERENCE: 12557-015001
; CURRENT APPLICATION NUMBER: US/11/191,588
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/771,708
; PRIOR FILING DATE: 2004-02-04
; PRIOR APPLICATION NUMBER: US 60/444,771
; PRIOR FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1069
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-191-588-17

Query Match          9.4%; Score 84.5; DB 7; Length 1069;
Best Local Similarity 23.0%; Pred. No. 46;
Matches 51; Conservative 36; Mismatches 56; Indels 79; Gaps 13;

QY 6 DPARNTTVKEFILNKDTG-----EVSSELKPHRVTVTIQNGKMSSTIVSE----- 51
Db 574 DFKNLCTVKEI---DTGASAAANVPETK-HRVSGTVVEGKDSKSQLATKKVKKPTIKN 629
QY 52 -----DFILPV-----YKG-----ELEKGYQFDGWEISGPEGKKD----- 81
Db 630 TEHRRAPESVTPIGPPVEVKAEAIQTCINYEIGIKVQINNGEPFSG--VIFVKNKPTDTCRV 687
QY 82 -----AGVINLSKDTFIKPV-FKKIEEKKERENKPTFDVSKKKNPQVNHSQLNESH 133
Db 688 EVANSNAATLVGLPKDFGMRPISLDNIDNETGKNK-----TKKGE-----ETP 732

QY 134 RKEDLQREHSQKSDSTKDTATVLDKNNISKST-----TNN 171
Db 733 LKDEI---EBFRQKQAAEFRCDCGLVDLLNGTYKSTVVIQTNN 772

RESULT 35
US-11-196-400-3
; Sequence 3, Application US/11196400
; Publication No. US20050287166A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 2007/3USODIV
; CURRENT APPLICATION NUMBER: US/11/196,400
; CURRENT FILING DATE: 2005-08-04
; PRIOR APPLICATION NUMBER: US/09/742,096
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,462
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.  
OM protein - protein search, using sw model  
Run on: April 24, 2006, 14:50:52 ; Search time 18.2953 Seconds  
(without alignments)  
915.083 Million cell updates/sec  
Title: US-10-067-385-8\_COPY\_600\_773  
Perfect score: 897  
Sequence: 1 KIVVKDFARNTTVKKEFILNK.....ATVLDKNNISSKSTNNPNK 174  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : PIR 80:\*  
1: pirl:\*  
2: pirl2:\*  
3: pirl3:\*  
4: pirl4:\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	2140	2 F95074	serine proteinase,
2	894	99.7	2144	2 A97942	metalloproteinase,
3	114.5	12.8	1038	2 JC5497	claustrin - chicke
4	110	12.3	558	2 T18467	hypothetical prote
5	109	12.2	665	2 B71609	hypothetical prote
6	104	11.6	540	2 D86432	hypothetical prote
7	103.5	11.5	325	2 T18283	hypothetical prote
8	102.5	11.4	385	2 T20410	hypothetical prote
9	102	11.4	622	2 A90570	lipoprotein (impor
10	101.5	11.3	312	2 G81339	probable membrane
11	101.5	11.3	535	2 T37189	hypothetical prote
12	101	11.3	700	2 S67610	probable membrane
13	100	11.1	219	2 B72291	hypothetical prote
14	100	11.1	1875	2 S38173	myosin-like protei
15	99	11.0	211	2 T25911	hypothetical prote
16	99	11.0	1397	2 T10466	DNA topoisomerase
17	97.5	10.9	614	2 A84152	hypothetical prote
18	97.5	10.9	1345	2 S46817	hypothetical prote
19	97	10.8	644	2 T47835	hypothetical prote
20	97	10.8	988	2 T14188	hypothetical prote
21	96.5	10.8	456	2 T05612	hypothetical prote
22	96.5	10.8	867	2 T27136	hypothetical prote
23	96.5	10.8	871	2 T27135	hypothetical prote
24	96	10.7	491	2 C97267	hypothetical prote
25	96	10.7	1939	2 T18372	repeat organellar
26	96	10.7	2500	2 G71609	hypothetical prote
27	95.5	10.6	3724	2 T18427	hypothetical prote
28	95	10.6	348	2 T37271	cylicin II - human
29	95	10.6	1202	1 S05362	probable DNA-dirc

30 94.5 10.5 277 2 D70214 surface lipoprotei  
31 94.5 10.5 670 2 T28391 ORF MSV230 hypothe  
32 94 10.5 210 2 T28771 hypothetrical prote  
33 94 10.5 553 2 T15094 hypothetrical prote  
34 94 10.5 947 2 T03795 ornithine decarbox  
35 93.5 10.4 456 2 T03045 hypothetrical prote  
36 93.5 10.4 645 2 E89883 conserved hypothet  
37 93.5 10.4 649 2 S42488 dnak-type molecula  
38 93 10.4 629 2 G96542 hypothetrical prote  
39 93 10.4 2464 1 QRMSP1 microtubule-associ  
40 92.5 10.3 443 2 S66040 serine-type D-Ala-  
41 92.5 10.3 571 2 B64469 hypothetrical prote  
42 92.5 10.3 1702 2 A41859 Iga-specific metal  
43 92.5 10.3 1888 2 T39009 hypothetrical prote  
44 92.5 10.3 1997 2 F71607 DNA helicase II BR  
45 92 10.3 688 2 A47705 triacylglycerol li

ALIGNMENTS

RESULT 1  
P95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: P95074  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel, non, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison A:Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae. A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: P95074  
A>Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-2140 <KUR>  
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match 100.0%; Score 897; DB 2; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.5e-56;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKG 60  
Db 1933 KIVVKDFARNTTVKKEFILNKDTGEVSELKPHRVTTVIQNGKMSSTIVSEEDFILPVYKG 1992  
Qy 61 ELEKGYQFDGWEISGFEGKDGAGVYVNLKDTFKVPFKIEKKKEENKPTFDVSKKD 120  
Db 1993 ELEKGYQFDGWEISGFEGKDGAGVYVNLKDTFKVPFKIEKKKEENKPTFDVSKKD 2052  
Qy 121 NPQVNHSQLNSHRKEDIQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 174  
Db 2053 NPQVNHSQLNSHRKEDIQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2106

RESULT 2  
A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942  
R:Roskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A97942  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KUR>  
A:Cross-references: UNIPROT:Q8DQ7; UNIPARC:UPI00000E3490; GB:AE007317; PIDN:AAK99365.1;  
C:Genetics:  
A:Gene: pta  
C:Keywords: hydrolase; serine proteinase

Query Match 99.7%; Score 894; DB 2; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 4.1e-56;  
Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLVVQDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 60  
DB 1937 KLVVQDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKG 1996

QY 61 ELEKGYQPDGWEISGPEGKDDAGYVNLKDTFIKVPFKIEEKEEENKPTFDVSKKD 120  
DB 1997 ELEKGYQPDGWEISGPEGKDDAGYVNLKDTFIKVPFKIEEKEEENKPTFDVSKKD 2056

QY 121 NQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 174  
DB 2057 NQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2110

RESULT 3  
JC5497  
Clausstrin - chicken  
N:Alternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence revision 12-Sep-1997 #text\_change 09-Jul-2004  
C:Accession: JC5497; PC4334; S37561  
R:Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Clausstrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JC5497; MUID:94157526; PMID:7906711  
A:Accession: JC5497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BURL>  
A:Cross-references: UNIPROT:Q90784; UNIPARC:UPI00000FC87F; EMBL:X67778; NID:g406318; PID  
A:Accession: PC4334  
A:Molecule type: protein  
A:Residues: 79-83;299-412;485-502 <BUR2>  
A:Cross-references: UNIPARC:UPI000017BFF3; UNIPARC:UPI000017BFF4; UNIPARC:UPI000017BFF5  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervo  
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,490/Binding site: carbohydrate (Aen) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 12.8%; Score 114.5; DB 2; Length 1038;  
Best Local Similarity 23.5%; Pred. No. 1.4;  
Matches 47; Conservative 29; Mismatches 65; Indels 59; Gaps 6;

QY 2 IVVKDFARN---TTVKEFILNKDTGEVSELKPHRVTVTIQ----- 39  
DB 514 VTQKDLTGNIASPAVQAQKLKQRTDSKSLKPAKTTTKDQCQKRLKKKHSLSLQVQ 573

QY 40 -----GKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGK 79  
DB 574 LEKPKLESEKPTPVKKEKAVRPETITVIAEKDV-----TTKEEQLGKSETSEKQAS 625

QY 80 KDAGYVINLSKDTFIKVPFK-KIEEKEEENKPTFDVSKKDNPNQVNHSQLNESHKEDL 138  
DB 626 EKQDVKPKVTYKESVKKEKAPKEEKKDEKKEPKKEVSKKEKPLI---KKEEKPKCEDI 682

QY 139 QRE-----EHSQKSDTKDV 153  
DB 683 KKEVKKEVKKEEKEAKEV 702

RESULT 4  
TJ18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: TJ18467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: TJ18467  
A:Status: preliminary; translated from (GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: UNIPROT:O77355; UNIPARC:UPI000017CC2E; EMBL:AL008970; NID:e1407852;  
C:Genetics:  
A:Map position: 3  
A:Introns: 84/1, 160/1  
A:Note: C0465C

Query Match 12.3%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 1.5;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 47 IVSEEDFILPVY-----KGELEKGYQPDGWEISGPEGK---DAGYVINLSKDTFIKVP 97  
DB 60 ILGFEDDILYEYCISQLKQSKK--KADGEEDKYLNAKKLAKINLTGPIGNKKSDFIEEL 117

QY 98 PKKI--BEKKE-----ENKPTFDVSK-KQNPQVNHSQLNE-----SHRK 135  
DB 118 LELLNBEKKEHIAIDLNLNENK-TNDIKKVNENINENYVNNENKOTISNKDKKEHVSQN 176

QY 136 E-----DLQREH-----SQKSDSTK---DVTATVLDKNNISSKSTTN 170  
DB 177 EHNINNVNLKKEKEYTDIQRDKRKHGSLSQKSDSYKRPFNKRTKSTIER-SLSNKRIDE 235

QY 171 NPNK 174  
DB 236 KTNK 239

RESULT 5  
B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: B71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Pertes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-665 <GAR>  
A:Cross-references: UNIPROT:O96229; UNIPARC:UPI000017B60A; GB:AE001410; GB:AE001362; NII  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0680w

Query Match 12.2%; Score 109; DB 2; Length 665;  
Best Local Similarity 26.1%; Pred. No. 2.1;  
Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;

QY 5 KDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELE 63  
DB 114 KDDNNNNNGTKQIEKKNKINKSDL--HRQNELNLQSGK-----NEQDI-----NKNE 158

QY 64 KGYQPDGWEI--SGPEGKDDAGYVINLSKDTFIKVPFKIEEKE-----EENKP 111  
DB 159 KGKQ-----DIGNSNAENKKD-----VKEGVKELEBEKKKEBKISDDHKVEENKK 202

QY 112 TFD-----VSKKKNPNQVNHSQLNESHKEDLQR-EEHQKSDSTKDVATVLDKNNISSK 166

RESULT 8  
T20410  
hypothetical protein E02A10.2 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T20410  
R;Thomas, K.  
submitted to the EMBL Data Library, October 1996  
A;Reference number: Z19271  
A;Accession: T20410  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-385 <WIL>  
A;Cross-references: UNIPROT:Q93424; UNIPARC:UPI000008935C8; EMBL:Z81053; PIDDN:C  
A;Experimental source: clone E02A10  
C;Genetics:

[illegible]

```

Db      317  -KKBEKKBEQKE--EVEKKEE-----EEKKDEPKCKBEKKBEKKEDRVEKSEKV 368
QY      158  LDK 160
        :|
Db      369  EEK 371

RESULT 9
A90570
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CT1P)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C:Accession: A90570
R;Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galissou, P.; M
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycop
A:Reference number: A90512; MUID:21267165; PMID:11353084

```

A;Accession: A90570  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-622 <KUR>  
A;Cross-references: UNIPROT:Q980A1; UNIPARC:UPI00000C80A1; GB:AL445566; PID:g1  
A;Experimental source: strain UAB CT1P

C;Genetics:  
A;Gene: MYPV 4650  
A;Genetic code: SGC3

Query Match 11.4%; Score 102; DB 2; Length 622;  
Best Local Similarity 24.9%; Pred. No. 6.3;  
Matches 48; Conservative 35; Mismatches 66; Indels 44; Gaps 11;

QY 9 RNTVKEPTL-NKOTGEVSEL-----KPHRV--TVTIQNG--KEM--SSTI--V 48  
DB 26 KNDKSAKFLKSSDLISISEKQFQFKNNLDRKNQKLIIEGVITFINSETKBIIVKKEITLNL 85  
QY 49 SBEEDFILPVYKGELEKGYQFDGWEISGPEGKDGAVVNLSDTKPIKVPFKKIEKKEE 108  
DB 86 SEDNIIFSLNLIENNAKFOLDPE-----VSKDEKFKLQFQEIINSQYEQ 129  
QY 109 NKPTFDVSKKD-----NPQVNHSQLNES--HRKEDLQREHSQKSDSTKVATVLDKRN- 161  
DB 130 -KITONISSKEDEKKNPKDNNSNNSSDQKNDLQKNNSDKLNDNVQDEKANKENSNS 188  
QY 162 NISSKSTTNPNK 174  
DB 189 NDSKEKNDENTNK 201

RESULT 10  
G81339  
probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: G81339  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: A81250; MUID:20150912; PMID:10689204  
A;Accession: G81339  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-312 <PAR>  
A;Cross-references: UNIPROT:Q9PPL5; UNIPARC:UPI00000C1CF0; GB:AL139076; GB:AL111168; NID  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj0692c

Query Match 11.3%; Score 101.5; DB 2; Length 312;  
Best Local Similarity 25.0%; Pred. No. 3.1;  
Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

QY 15 ERTLNKDTGEVSELKPHRVTVTIQNGKEMSTI---VSEEDFILPVYK-----GELE 63  
DB 5 DFIKERQNTIRQKMLKFSR---AINQKPLDDLDLDEISSDDILRRFPKKTTPNKPLEELD 61  
QY 64 KGYPQDWEISGPEGKDGAVVNL---SKDTPIKVPFKKIEKKEENKPT---FDVSK 117  
DB 62 BEVESKHTKSNLYLKED---LINVKLEKQSLAKKIPSKMKERRKEENKTKKNFLFSR 118  
QY 118 KQNP-----QVNHSQLNESHRKEDLQREHSQKSDSTKV--TATVLDKQNSSK--STT 169  
DB 119 KKANEIKNTQTQIQTKSNQATTQTKQEKELTNSIEKIQTETKIQKPLIEKLDVYK 178  
QY 170 NNPN 173  
DB 179 NQPN 182

RESULT 11  
T37189  
hypothetical protein C02H7.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T37189  
R;Leimbac, D.; Minx, M.

submitted to the EMBL Data Library, February 1996  
A;Description: The sequence of C. elegans cosmid C02H7.  
A;Reference number: Z20523  
A;Accession: T37189  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-535 <LRI>  
A;Cross-references: UNIPROT:Q17595; UNIPARC:UPI000008019E; EMBL:U49945; PIDN:AAC47924.1;  
A;Experimental source: strain Bristol N2; clone C02H7  
C;Genetics:  
A;Gene: CESP:C02H7.1  
A;Map position: X  
A;Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 11.3%; Score 101.5; DB 2; Length 535;  
Best Local Similarity 21.0%; Pred. No. 5.7;  
Matches 39; Conservative 31; Mismatches 71; Indels 45; Gaps 6;

QY 4 VKDFARNTTVKBEFILNK-----DTGRVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVY 58  
DB 55 IKSAGNDKNTKTAFLDKLILKLDGSLKNVKAAIL---SGDAEET----- 98  
QY 59 KGELEKGYQFDGWEISGPEGKDGAGVNLNSKOTFIKVPFKKIEKKEENKPTFDVSKK 118  
DB 99 ----NQMLQMLGTNATSFNSRNGTG-----BEKKKKKKVKCKDKGDEEEKST---TKK 144  
QY 119 KQNPQVNHSQLNESHRKEDLQREHSQKSDTK-----DVTATVLDKNNISSKST 168  
DB 145 RSSKKEETHEKEKKEKKSAEKEKKKKSSSKSRHKSRSRSEKSSKEKKEKKEKST 204  
QY 169 TNNPNK 174  
DB 205 TDEKPK 210

RESULT 12  
S67610  
probable membrane protein YDL074c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein D2483  
C;Species: Saccharomyces cerevisiae  
C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004  
C;Accession: S67610  
R;Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.  
submitted to the Protein Sequence Database, July 1996  
A;Reference number: S67608  
A;Accession: S67610  
A;Molecule type: DNA  
A;Residues: 1-700 <WAM>  
A;Cross-references: UNIPROT:Q07457; UNIPARC:UPI0000069BFF; EMBL:Z74122; NID:gl431087; P1  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:BREL; MIPS:YDL074C  
A;Cross-references: SGD:S0002232  
A;Map position: 4L  
C;Keywords: transmembrane protein  
P;69-85/Domain: transmembrane #status predicted <TM>

Query Match 11.3%; Score 101; DB 2; Length 700;  
Best Local Similarity 22.4%; Pred. No. 8.5;  
Matches 46; Conservative 39; Mismatches 60; Indels 60; Gaps 10;

QY 5 KDFARNTTVKBEFILNK-----DTGEVSELKPHRVTVTIQ-----NGKEMST 46  
DB 379 KQVRIARTARDLLSKTAILEAKSKTEVLSDLQ-HAIDILKEQWKIDQSRNDTKSSST 437  
QY 47 IVSEEDFILPVYKGELEKGYQFDGWEISGFGKK-----DAGYV 85  
DB 438 ----QDALIKEIQ-DLEKGR----ELSDLTHKKYSEIINHESVLSKLTVEKTKADQKYF 488  
QY 86 INL-SKDTF---IKVPFKKIEKKEENKPTFDVSKKQNPQVNHSQLNESHRKEDLQRE 141  
DB 489 AMRSKDSILTEIKTLTKSLSKSNEL-----ILQLKSDRLLOOKIGNLHKQJDLSON 541



**RESULT 13**

Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing of *Methanocaldococcus jannaschii*  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: S72291  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-219 <ARN>  
A;Cross-references: UNIPROT:Q9X0M6; UNIPARC:UPT00000C12ED; GB:AE001771; GB:AE000512; NID:10360571  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TW1142

DB 816 LSELKKET'SQKDHHIKQLEEDNNSN 840

	Query Match	11.0%;	Score 99;	DB 2;	Length 211;
	Best Local Similarity	29.4%;	Pred. No. 3;		
	Matches	30;	Conservative 17;	Mismatches 51;	Indels 4; Gaps 1;
Gy		77	EKKDAGYVINLSKPTFKPVKCKIEEKKBEENKNPFTFVSKGKDNPQNVNH----	SOLNES	132
Dg		69	EGEKGDEKKSGDKGDKEEKDDSEKGGDGCGEDDKDEKKDDKDEKKDADEKKDE		128
Gy		133	HAKEDIQRHEHSGKS DSTT KDVTATVLVDKNNIS KSTTNPNPK		174
Dg		129	EKKDDPKCDKDEKKDDEKKDEKKESKSSKGS KKSKS K		170

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium falciparum)  
C/Species: Plasmodium falciparum  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C/Accession: T10466  
R/Cheesman, S.J.  
Submitted to the EMBL Data Library, September 1995  
A/Reference number: Z17031  
A/Accession: T10466  
A/Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-1397 <CHR>  
A:Cross-references: UNIPARC:UPI00000006D7; EMBL:X79345; NID:g994807  
C:Genetics:  
A:Gene: TopoII  
A:Map position: 14  
C:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hyd  
C:Keywords: ATP; DNA binding; isomerase; nucleus

Query Match 11.0%; Score 99; DB 2; Length 1397;  
Best Local Similarity 22.2%; Pred. No. 26;  
Matches 44; Conservative 45; Mismatches 67; Indels 42; Gaps 9;

QY 1 KIVVKDFAR-----NTTVKEPILNKDTEVSELKPHRVTVTIQNGKMSSTIVSEE 51  
DB 1092 KVLVELYKRGVDPYKINDINKKBEIFQELLDAAD-NPE-----DNEEIIAGITVKDY 1144  
QY 52 DFIL--PVYKBLEKGYQPDGWEISGFEGKDGAGVINLSKDTFIKVPFKIEE----- 103  
DB 1145 DYLLSWPIFSLTLEK---VEDLLTQLKEKERLEILRNITVETMWLKDIEKVEEAIEFOR 1201  
QY 104 -----KKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQ---KSDSTKDVTA 155  
DB 1202 NVELSNREESNK--FKVARKQ-----GPSMMKKKKKKKLLSSDSESEGGDTSDSSEFLVN 1254  
QY 156 TVLDKNNISSKSTNNPN 173  
DB 1255 TLNIKNTWKTKITSSNN 1272

RESULT 17  
A84152  
hypothetical protein BH4017 [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans  
C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C:Accession: A84152  
R:Takami, H.; Nakagawa, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
A:Reference number: A83650; MUID:20512582; PMID:11059132  
A:Accession: A84152  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-614 <STO>  
A:Cross-references: UNIPROT:Q9K5S1; UNIPARC:UPI0000004396; GB:AP001520; GB:BA0000004; NID  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: BH4017

Query Match 10.9%; Score 97.5; DB 2; Length 614;  
Best Local Similarity 27.3%; Pred. No. 13;  
Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;

QY 26 SELKPHRVTVTIQNGKMSSTIVSE-----EDFILPVYKBLEKGYQPDGW-----EISG 75  
DB 386 SEPKPE--TYTLQTAIQM-TPIVNEYSPTREFL-----ARKAQLDGHADVSKVTY 435  
QY 76 FEGKKDAGVINLSKDTFIKVPFKIEEK--EENKPTFDVSKK---DNPQVNHSQL 129  
DB 436 FAGRNIASQL-----GKIEEKLQDKYNNYTFDFPKVEVNVQSPIKSTSA 482  
QY 130 NESHKEDLQREHSOKSDSTKDVTAIVLD 159  
DB 483 GKVGIVGLGRKYHKGEDLERRLSKSQIE 512

RESULT 18  
S46817  
hypothetical protein YHR080c - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Oct-2004  
C:Accession: S46817  
R:Favetto, T.

submitted to the EMBL Data Library, June 1994  
A:Description: The sequence of S. cerevisiae cosmid 9205.  
A:Reference number: S46795  
A:Accession: S46817  
A:Molecule type: DNA  
A:Residues: 1-1345 <PAV>  
A:Cross-references: UNIPROT:P38800; UNIPARC:UPI000013B2B1; EMBL:U10556; NID:g500825; PID  
C:Genetics:  
A:Cross-references: SGD:S0001122  
A:Map position: 8R  
C:Superfamily: uncharacterized conserved protein  
C:Keywords: transmembrane protein

Query Match 10.9%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 32;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 59 KGELEKGYQPDGWEISGFEGKK-DAGVYVINLSKDTFIKVPFKIEEKKKEENKPTFDVSK 117  
DB 1109 KGAIEKG-----SVEGQKVSQYMLSELRII-----SRASKKPVKVKWK 1149  
QY 118 KQNPQVNHSQLNESHKEDLQREHSOKSDSTKDVTAIVLD--KNNISSKSTTN---NP 172  
DB 1150 SHDKHRPFPKVE-----QKSSERKSDDNKDILTHILDFVQNNFSSFIWMKLLSP 1201  
QY 173 NK 174  
DB 1202 QK 1203

RESULT 19  
T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C:Accession: T47835  
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: 224475  
A:Accession: T47835  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <NVA>  
A:Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI000009E0AD; EMBL:AL138658  
A:Experimental source: cultivar Columbia; BAC clone T209  
C:Genetics:  
A:Map position: 3  
A:Introns: 158/2; 329/3  
A:Note: T209.90

Query Match 10.8%; Score 97; DB 2; Length 644;  
Best Local Similarity 22.9%; Pred. No. 15;  
Matches 43; Conservative 29; Mismatches 72; Indels 44; Gaps 7;

QY 8 ARNTTVKEFILNKO--TGEV--SELKPHRVTVT-----IQNGKMSSTIVSEE--- 51  
DB 476 SRRSKEIAVAVAKDTKGRKNIKKQDTTKTESDDDDDEKEENSKTEKKTADVADKKSV 535  
QY 52 -DFILPVYKBLEKGYQPDGWEISGFEGKKDAGVYVINLSKDTFIKVPFKIEEKKKEENK 110  
DB 536 ADFLKRIKKNSPQKGGK-----ETTSXNQKQKDGNV-----KKENDHQ 572  
QY 111 PTFVSKKKONPQVNHSQLNESHKEDLQREHSOKSDS---TKDVTATVLDKNNISSK 166  
DB 573 KKSQGNVKKSKVKPRRLSRSTGRKRVENNNNSKSKSRKKQTKETAIVATGKRGRESG 632  
QY 167 STTNPNK 174  
DB 633 KDDKQPRK 640

RESULT 20  
T14188

hypothetical protein T28D5.30 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T14188  
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: Z17931  
A:Accession: T14188  
A:Molecule type: DNA  
A:Residues: 1-988 <BEV>  
A:Cross-references: UNIPROT:Q9STN4; UNIPARC:UPI00000A3618; EMBL:AL109819  
A:Experimental source: cultivar Columbia; BAC clone T28D5  
C:Genetics:  
A:Gene: ATSP-T28D5.30  
A:Map position: 4  
A:Introns: 162/3; 201/3; 416/3; 438/3; 460/3; 482/3; 504/3; 519/3; 534/3; 559/3; 579/3;  
C:Superfamily: Arabidopsis thaliana hypothetical protein T21C14.40

Query Match 10.8%; Score 97; DB 2; Length 988;  
Best Local Similarity 22.1%; Pred. No. 24;  
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;

QY 39 NGKEMSTIVSEDFILPVYKGLKGY-----QPDGWEISGPEGKDGAGYVNLNLSKDTF 93  
DB 383 NGRQNSNVQSSVDEILSYTKVPSGVGLNVSERDVELVEDDVRSAAGLSPVQRDN- 441

QY 94 IKPVFKKIBKKE-----ENKPTFDVSKKKNPQVNHSQLNESHKREDLQREE 142  
DB 442 VEPVGDDVRSSGDMSPNAAANVRGPPATFDIMESDNPGRDNVAPMEDHRSVQLSP 501

QY 143 HSQKSDSTKDT--ATVLDKNNISSKSTNNPNK 174  
DB 502 HVL--GAKDVTDSPTDKGVNDVTDASDPT 532

RESULT 21  
T05612  
hypothetical protein F9D16.270 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: T05612  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05612  
A:Molecule type: DNA  
A:Residues: 1-456 <BEV>  
A:Cross-references: UNIPROT:Q9SUP7; UNIPARC:UPI00000A81F5; EMBL:AL035394  
A:Experimental source: cultivar Columbia; BAC clone F9D16  
C:Genetics:  
A:Map position: 4  
A:Introns: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
A:Note: F9D16.270

Query Match 10.8%; Score 96.5; DB 2; Length 456;  
Best Local Similarity 25.0%; Pred. No. 11;  
Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;

QY 20 KDTGEVSEKLP-HRYVTVTQNGKEMSSIVSEDFILPVYKGELEKGYQPDGWEISGREG 78  
DB 245 KKKGKQDPLKPKHPVSFAFLVYANERRAALREENKSGVVEAK-----ITGEEWKNLSD 296

QY 79 KKDAGY--VINLSKDTFIKVPVKIIEKKEENKPTFDVSKKKNPQVNHSQLNESHKRE 136  
DB 297 KKKAPYKAKKKETYLQ-AMEBYKRYKEE-----ALSQKE-----EEELLKHKQ 345

QY 137 DLQREHSQKSDSKDVTATVLDKNNISSKSTNN--PNK 174  
DB 346 ALQMLKKKEKTDN-----LIKKEKATKKKNVNDPNK 378

RESULT 22  
T27136

hypothetical protein Y53C12B.3a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T27136  
R:Kershaw, J.; Lennard, N.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20316  
A:Accession: T27136  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-867 <WIL>  
A:Cross-references: UNIPROT:O18218; UNIPARC:UPI0000075065; EMBL:Z99278; PIDN:CAB16493.  
A:Experimental source: clone Y53C12B  
C:Genetics:  
A:Gene: CESP:Y53C12B.3a  
A:Map position: 2  
A:Introns: 100/3; 177/3; 218/1; 423/3; 714/2; 864/3

Query Match 10.8%; Score 96.5; DB 2; Length 867;  
Best Local Similarity 25.0%; Pred. No. 23;  
Matches 45; Conservative 33; Mismatches 63; Indels 39; Gaps 8;

QY 9 RNTTVKEFILNKDTG-----EVSELKPHRYVTVTQNGKEMSSIVSEDFILPVYKGL 62  
DB 651 KKTPKCKDPKVDQEASIEATKTEISEENPK--TDDIQSKDDVTS-----KSEL 695

QY 63 E----KGQFDGWEISGPEGKDGAGYVNLNLSKDTFIKVPVKIE---EKKEENKPTPD 114  
DB 696 HCYRCEHYQLPABEVSSHNIRKNG-----DLWCEHM-KKIKGCHCEATGEGHHPLI 747

QY 115 VSKKKNPQVNHSQLNESHKREDLQREHSQKSDSKDVTATVLDKNNISSKSTNNPNK 174  
DB 748 CPKKEERVAKSR--ESSQKPIDPQEISDDQDDTVPDPQIVQDNQSHKSHNSNR 805

RESULT 23  
T27135  
hypothetical protein Y53C12B.3b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T27135  
R:Kershaw, J.; Lennard, N.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20316  
A:Accession: T27135  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-871 <WIL>  
A:Cross-references: UNIPROT:O62486; UNIPARC:UPI0000079598; EMBL:Z99278; PIDN:CAB16492.1  
A:Experimental source: clone Y53C12B  
C:Genetics:  
A:Gene: CESP:Y53C12B.3b  
A:Map position: 2  
A:Introns: 100/3; 177/3; 218/1; 423/3; 717/2; 867/3

Query Match 10.8%; Score 96.5; DB 2; Length 871;  
Best Local Similarity 25.0%; Pred. No. 23;  
Matches 45; Conservative 33; Mismatches 63; Indels 39; Gaps 8;

QY 9 RNTTVKEFILNKDTG-----EVSELKPHRYVTVTQNGKEMSSIVSEDFILPVYKGL 62  
DB 654 KKTPKCKDPKVDQEASIEATKTEISEENPK--TDDIQSKDDVTS-----KSEL 698

QY 63 E----KGQFDGWEISGPEGKDGAGYVNLNLSKDTFIKVPVKIE---EKKEENKPTPD 114  
DB 699 HCYRCEHYQLPABEVSSHNIRKNG-----DLWCEHM-KKIKGCHCEATGEGHHPLI 750

QY 115 VSKKKNPQVNHSQLNESHKREDLQREHSQKSDSKDVTATVLDKNNISSKSTNNPNK 174  
DB 751 CPKKEERVAKSR--ESSQKPIDPQEISDDQDDTVPDPQIVQDNQSHKSHNSNR 808

RESULT 24







Query Match 10.4%; Score 93.5; DB 2; Length 456;  
Best Local Similarity 20.3%; Pred. No. 18;  
Matches 41; Conservative 39; Mismatches 83; Indels 39; Gaps 5;  
QY 3 VVKDPAINTVTVKGFILNKDTGVSELKPHRVTVTTIONGKEMSTIVSEEDFILPVYKGL 62  
DB 70 VVQSISINTPIETI--KDTPIYEE--TPIKRTITVNTVKSSISGNGRNLDDDL 126  
QY 63 EKGYPQDWEISGFEKGDAGVYVNLKDTPI-----RPVFKKIBKKKE-----BENKPT 112  
DB 127 PDDRYKSPTRKFGQEKDEDIRLIPKSSNIGSKYKPVLRVVENENKKIHIQKES 186  
QY 113 FVUSKKKNQ-----VNHSQLNESHKEDLQREHSQK 146  
DB 187 IVNDERKKNPFRKPKDKNEDKKVPPPSLKEIENKGDHEENEDKKELMFKLQLLQK 246  
QY 147 SDSTKDVATVLDKNNISSKST 168  
DB 247 QYPLRDIPDFTIRSEYKMKKT 268

## RESULT 36

E89883 conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: E89883  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.; Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: E89883  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-645 <KUR>  
A:CROSS-references: UNIPROT:Q99UX5; UNIPARC:UPI00000CAAB3; GB:BA000018; PID:gl3700929; E89883  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0976

Query Match 10.4%; Score 93.5; DB 2; Length 645;  
Best Local Similarity 21.9%; Pred. No. 27;  
Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;  
QY 11 TTVKGFILNKDTGE-VSELKPHRVTV--TIQNGKEMSTIVSEEDFILPVYKGELE-KGY 66  
DB 327 SAITEFQNVQPTNEKMTLDQTKYVYVESVENNESMMDTFVKH-----PIKTMGLNGKKY 381  
QY 67 QP-----DGEISGFEGKK-----D 81  
DB 382 MWMTETNDYDKDFWVGQVRVTSKDAKNTRTILPPVVEKTYLDALVKVHTIIDYD 441  
QY 82 AGYVNL-SKDTFIKPVFKKIBKKKEENKPTFDV-----SKKDNQPNVNHSQLNESHK 135  
DB 442 GQYHVRVDKEAFTKANTDKSNKQODNSAKKEATPATPKSPTEPSPEKESQKDSQKD 501  
QY 136 EDLQ-----REHSQKSDSTKQVT-ATVLDKNNISSKSTNNPNK 174  
DB 502 DNKQLPSVEKENDASSEGKDKTPATKPTKGEVSESSSTT--PTK 543

## RESULT 37

S42486 dnaK-type molecular chaperone hsp70 - Pyrenomonas salina nucleomorph  
N:Alternate names: heat shock protein 70  
C:Species: nucleomorph Pyrenomonas salina  
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C:Date: 13-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 31-Dec-2004  
C:Accession: S42488; S45576  
R:Resing, S.; Hofmann, C.J.B.

submitted to the EMBL Data Library, March 1993  
A:Description: Smallest known eukaryotic genomes encode a protein gene: towards undere.  
A:Reference number: S42488  
A:Accession: S42488  
A:Molecule type: DNA  
A:Residues: 1-649 <RES>  
A:CROSS-references: UNIPROT:P37899; UNIPARC:UPI000012CC4F; EMBL:X72621; NID:g461335; P.  
R.Hofmann, C.J.B.; Resing, S.A.; Haeuber, M.M.; Martin, W.F.; Mueller, S.B.; Couch, J.  
Mol. Gen. Genet. 243, 600-604, 1994  
A:Title: The smallest known eukaryotic genomes encode a protein gene: towards an under  
A:Reference number: S45576; MUID:94268506; PMID:8208251  
A:Accession: S45576  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-24,'Q',26-91,'H',93-578 <HOF>  
A:CROSS-references: UNIPARC:UPI0000177D1F; EMBL:X72621; NID:g461335  
A:Note: the sequence is revised in GenBank entry PSHSP70, release 111.0, (PIDN:CAA5119  
C:Genetics:  
A:Gene: hsp70  
A:Genome: nucleomorph  
C:Function:  
A:Description: involved in protein folding and assembling/diseassembling of protein com  
C:Superfamily: bcr protein  
C:Keywords: ATP; molecular chaperone; nucleomorph  
Query Match 10.4%; Score 93.5; DB 2; Length 649;  
Best Local Similarity 24.5%; Pred. No. 27;  
Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;  
QY 17 ILNKDTGVSELKPHRVTVTTIONG---KEMSTIVSEEDFILPVYKGELEKGYQDQGEI 73  
DB 491 ILNVSADSKSTGKGNKITITNDKRLSKETRMVVEAB-----KYTKDEK----- 537  
QY 74 SGFEKKDA-----GYVINLS---KDTPIKPVFKKIBKKKEENKPTFDVSKKDNQPN 125  
DB 538 --LDKLEAKNSLENYAIVRTVRD-----EKLKIQIEDKKSIEEKVK----- 582  
QY 126 HSQNLNESHKEDLQREHSQKSDSTKQVTVATVLDK 160  
DB 583 --VLEFIETNEDLEKEEYEEKEKELKNPANDPIISK 615  
RESULT 38  
G96542  
Hypothetical protein F1736.14 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: G96542  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A96141; MUID:21016719; PMID:11130712  
A:Accession: G96542  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-629 <STO>  
A:CROSS-references: UNIPROT:Q9C6P8; UNIPARC:UPI00000A743C; GB:AB005173; NID:gl11054631;  
C:Genetics:  
A:Gene: F1736.14  
A:Map position: 1  
Query Match 10.4%; Score 93; DB 2; Length 629;  
Best Local Similarity 19.6%; Pred. No. 28;  
Matches 40; Conservative 39; Mismatches 73; Indels 52; Gaps 7;  
QY 11 TTVKGFILNKDTGEV-----SELKPHRVTVTTIONGKEMSTIVSEEDFI 54

Db	391	TTAK---MSSSTAELVQLPAEKGVGMKDQKVQSQEGPHLETAKPTKDSAMEQTVAEADVAM	447
Qy	55	LPVYGELEKGYQFDGWEISFGPEGKDGAGVYNLSKDTPIKVPVFKKIEKKE	106
Db	448	NPIVEKAMSEVAEAGAAINPIVEAEDG-----AMNPIVEKAMSGQVIEAEDAALNQAVD	501
Qy	107	---BENKPTFDVSKKKNQ---VNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKN	161
Db	502	ANFQTAPTGNDADSDDPSEPVSHS---ETLNPPELEKKEVYMRKDATERSVSADCDQKN	558
Qy	162	N-----ISSKSTTNNPNK	174
Db	559	SKLIAESSLQEISASQTSNSTQ	582
RESULT 39			
QRMSP1			
Microtubule-associated protein MAP1B - mouse			
N/Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei			
C/Species: Mus musculus (house mouse)			
C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Jul-2004			
C/Accession: S07549; S44387; A33645			
R/Noble, M.; Lewis, S.A.; Cowan, N.J.			
J. Cell Biol. 109, 3367-3376, 1989			
A/Title: The microtubule binding domain of microtubule-associated protein MAP1B contains			
A/Reference number: A33645; MUID:190094539; PMID:2480963			
A/Accession: S07549			
A/Molecule type: mRNA			
A/Residues: 1-2464 <NOB>			
A/Cross-references: UNIPROT:P14873; UNIPARC:UPI0000297D7; EMBL:X51396; NID:G52999; PIDN			
R/Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.			
Arch. Biochem. Biophys. 310, 428-432, 1994			
A/Title: Binding of heat-shock protein 70 (hsp70) to tubulin.			
A/Reference number: S44387; MUID:94234720; PMID:8179328			
A/Accession: S44387			
A/Status: preliminary			
A/Molecule type: protein			
A/Residues: 653-663, 'IC' <SAN>			
A/Cross-references: UNIPARC:UPI0000173097			
C/Superfamily: microtubule-associated protein MAP1B			
C/Keywords: microtubule binding; phosphoprotein; tandem repeat			
F/589-786/Domain: microtubule binding #status experimental <MTB>			
F/589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69			
R-K-E/D-X)			
F/1861-2064/Region: 17-residue repeats			
F/91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: ph			
F/147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (Co			
F/1953/Binding site: phosphate (Tyr) (covalent) #status predicted			
Query Match 10.4%; Score 93; DB 1; Length 2464;			
Best Local Similarity 27.8%; Pred. No. 1.3e+02;			
Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;			
Qy	14	KEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVVKGELEKGYQFDGWEI	73
Db	584	EKVLVKDKRPVTEKSP-----SVT-----EKEVSS---KEEQ--SPV-KAEVA-----EK	623
Qy	74	SGFEGKDGAGVYNLSKDTFIKVPFK-KLEEKKEBENKPTFDVSKKKNQPNVNHSQLNES	132
Db	624	QATESKP-----KVTQDKVVKKEIKTKLEEKKEE--KPKKEVVKEDKTPV---KKDEK	672
Qy	133	HKEDLQRE-----EHSQKSDSTKDVATVLDKNINSSKSTTNNPNK	174
Db	673	PRKEEVKKEIKKEIKKEERKELKGVKKEETPLKDAKGVKKEKVEKKEEPKK	728
RESULT 40			
S66040			
serine-type D-Ala-D-Ala carboxypeptidase (BC 3.4.16.4) dacA - Bacillus subtilis			
N/Alternate names: penicillin-binding protein 5			
C/Species: Bacillus subtilis			
C/Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004			

C/Accession: S66040; I39830; A92275; A92307; B61335; D69612; A23307	R/Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994	
A/Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom	A/Reference number: S65967; MUID:96051385; PMID:7584024
A/Accession: S66040	
A/Status: preliminary; nucleic acid sequence not shown; translation not shown	A/Molecule type: DNA
A/Residues: 1-443 <OG>	
A/Cross-references: UNIPROT:P08750; UNIPARC:UPI000005FDBA; EMBL:D26185; NID:G467326; PID	A/Note: The nucleotide sequence was submitted to the EMBL Data Library, December 1993
R/Todd, J.A.; Roberts, A.N.; Johnstone, K.; Piggot, P.J.; Winter, G.; Ellar, D.J.	
J. Bacteriol. 167, 257-264, 1986	A/Title: Reduced heat resistance of mutant spores after cloning and mutagenesis of the B
A/Reference number: I39830; MUID:86250602; PMID:3087956	
A/Accession: I39830	A/Status: preliminary; translated from GB/EMBL/DBDJB
A/Molecule type: DNA	
A/Residues: 48-226, 'O', 228-443 <RES>	A/Cross-references: UNIPARC:UPI000016E82C; GB:M13766; NID:G142816; PIDN:AAA22375.1; PID:
R/Waxman, D.J.; Strominger, J.L.	
J. Biol. Chem. 255, 3964-3976, 1980	A/Title: Sequence of active site peptides from the penicillin-sensitive D-alanine carbox
A/Reference number: A92275; MUID:80182289; PMID:6768745	
A/Accession: A92275	A/Molecule type: protein
A/Residues: 32-95, 'X', 97-98, 'XQX', 102 <WAX>	
A/Cross-references: UNIPARC:UPI000017804E	R/Waxman, D.J.; Strominger, J.L.
J. Biol. Chem. 256, 2067-2077, 1981	
A/Title: Primary structure of the COOH-terminal membranous segment of a penicillin-sensi	A/Reference number: A92307; MUID:81117303; PMID:6780559
A/Accession: A92307	
A/Molecule type: protein	A/Residues: 414-443 <WAX2>
A/Cross-references: UNIPARC:UPI000017804F	
R/Yocum, R.R.; Waxman, D.J.; Rasmussen, J.R.; Strominger, J.L.	Proc. Natl. Acad. Sci. U.S.A. 76, 2730-2734, 1979
A/Title: Mechanism of penicillin action: penicillin and substrate bind covalently to the	
A/Reference number: A61335; MUID:79223865; PMID:111240	A/Accession: B61335
A/Molecule type: protein	
A/Residues: 55-68 <YOC>	A/Cross-references: UNIPARC:UPI00000178050
R/Kunst, P.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter	
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho	Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler	
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.P.	Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel	
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, E.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle	Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror	
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,	T.; Winfers, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A/Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Banchin, A.	
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.	A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: D69612	
A/Status: nucleic acid sequence not shown; translation not shown	A/Molecule type: DNA
A/Residues: 1-443 <KUN>	
A/Cross-references: UNIPARC:UPI000005FDEA; GB:Z99104; GB:AL009126; NID:G2632267; PIDN:CA	A/Experimental source: strain 168
C/Genetics:	
A/Gene: dacA	A/Start codon: TTG
C/Superfamily: penicillin-binding protein 5	
C/Keywords: cell wall synthesis; hydrolase; serine carboxypeptidase	F/67/Active site: Ser #status experimental



Query Match	10.3%;	Score	92.5;	DB	2;	Length	443;
Best Local Similarity	25.0%;	Pred. No.	20;				
Matches	36;	Conservative	26;	Mismatches	49;	Indels	33;
						Gaps	6;
Qy	6	DEA-RNTTVKFEILNKDTGVSLSKPHRVTVTQNGKMSSTIVSSEDFILPVYKGELEK	64				
Db	304	YADPFIKSMKRIYAEGD-----QVGHK-TISVDGKKEKEVGIVTNKAPSLPVKNGE-EK	356				
Qy	65	DYQDFCWEISGFEGKKDAGVIVILNSKDTFTKPPFKKIEEKKEBENKPTFDVSKKQNPQV	124				
Db	357	NYRAK-----VTLNKDLTAPVKKGTKVCK-----LTAEYTGDEK	391				
Qy	125	NHSQLNESHKEDLQREHSQKSD	148				
Db	392	DYGLFNSDLGAVDLVTKNVEKAN	415				

Search completed: April 24, 2006, 15:01:18  
Job time : 20.2953 secs

**THIS PAGE BLANK (uspio)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 105.685 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897

Sequence: 1 KLVKDFARNTTVKPFILNK.....ATVLDKNISSKSTNNPNK 174

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	897	100.0	2119	2	Q9AHT5_STRPN	Q9Aht5 streptococc
2	897	100.0	2140	2	Q97RY6_STRPN	Q97ry6 streptococc
3	894	99.7	2144	2	Q8DQP7_STRPN	Q8dqp7 streptococc
4	893	99.6	2144	2	Q9S4M8_STRPN	Q9s4m8 streptococc
5	139	15.5	300	2	Q4XUI6_PLACH	Q4xui6 plasmodium
6	119	13.3	775	2	Q8CFK8_STRPN	Q8cpk8 staphylococ
7	117	13.0	1069	2	Q512T7_ENTHI	Q512t7 entamoeba h
8	115.5	12.9	296	2	Q50LX8_ENTHI	Q50lx8 entamoeba h
9	114.5	12.8	1038	2	Q90784_CHICK	Q90784 gallus gall
10	114	12.7	609	2	Q812K8_PLAF7	Q812k8 plasmodium
11	112.5	12.5	346	2	Q9U0G0_PLARE	Q9u0g0 plasmodium
12	112.5	12.5	354	2	Q2S995_PLAFA	Q2s995 plasmodium
13	112.5	12.5	354	2	Q81J55_PLAF7	Q81j55 plasmodium
14	112.5	12.5	361	2	Q9SP15_PLAFA	Q9sp15 plasmodium
15	112.5	12.5	379	2	Q9U6C4_PLAFA	Q9u6c4 plasmodium
16	112.5	12.5	662	2	Q4YMU4_PLABE	Q4ymu4 plasmodium
17	112	12.5	829	2	Q815P3_PLAF7	Q815f3 plasmodium
18	111.5	12.4	379	2	Q2S706_PLAFA	Q2s706 plasmodium
19	111.5	12.4	775	2	Q5HQ11_STABQ	Q5hq11 staphylococ
20	110.5	12.3	379	2	Q2S705_PLAFA	Q2s705 plasmodium
21	110	12.3	380	2	Q2G019_PLAFA	Q2g019 plasmodium
22	110	12.3	500	2	Q6BGL7_PASTE	Q6bgl7 paramesium
23	110	12.3	674	2	Q7BLE7_PLAYO	Q7ble7 plasmodium
24	109.5	12.2	384	2	Q50VJ0_ENTHI	Q50vj0 encamoeba h
25	109.5	12.2	3008	2	Q81436_PLAF7	Q81436 plasmodium
26	109	12.2	951	2	Q9G229_PLAF7	Q9g229 plasmodium
27	108.5	12.1	600	2	Q77355_PLAF7	Q77355 plasmodium
28	108.5	12.1	973	2	Q7RB37_PLAYO	Q7rb37 plasmodium
29	108	12.0	467	2	Q59PE2_CANAL	Q59pe2 candida alb
30	108	12.0	467	2	Q59PL2_CANAL	Q59pl2 candida alb
31	107.5	12.0	470	2	Q9FJK9_ARATH	Q9fjk9 arabidopsis

32	107.5	12.0	948	2	Q9U4U6_PLAFA	Q9u4u6 plasmodium
33	107.5	12.0	1455	2	Q640L5_MOUSE	Q640l5 mus musculu
34	106.5	11.9	374	2	Q5V9M0_PLAKN	Q5v9m0 plasmodium
35	106.5	11.9	1015	2	Q5W5T1_TETPY	Q5w5t1 tetrahymena
36	106	11.8	616	2	Q6BRW2_DEBHA	Q6brw2 debaromyce
37	106	11.8	988	2	Q7RP53_PLAYO	Q7rp53 plasmodium
38	106	11.8	1268	2	Q54HA7_DICDI	Q54ha7 dictyosteli
39	106	11.8	3026	2	Q81LS9_PLAF7	Q81ls9 plasmodium
40	105.5	11.8	329	2	Q9NFV9_PLAFA	Q9nfv9 plasmodium
41	105.5	11.8	605	2	Q7RJ33_PLAYO	Q7rj33 plasmodium
42	105.5	11.8	736	2	Q4VYV2_PLABE	Q4vyv2 plasmodium
43	105.5	11.8	2081	2	Q9LH98_ARATH	Q9lh98 arabidopsis
44	105	11.7	1550	2	Q54GS1_DICDI	Q54gs1 dictyosteli
45	105	11.7	3127	2	Q8IDA0_PLAF7	Q8ida0 plasmodium

#### ALIGNMENTS

RESULT 1

Q9AHT5\_STRPN  
ID Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5\_1  
DT 01-JUN-2001 (TRENBLrel. 17, Created)  
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
DOI=10.1128/IAI.69.3.1593-1598.2001;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
Langermann S., Johnson S., Koenig S.;  
RA "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08.064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0003086; F:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; TIGR01167; LPTXG anchor; 1.  
DR TIGRfams; TIGR01167; LPTXG anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Protease.

```

FT NON TER 1 1
SQ SEQUENCE 2119 AA; 238227 MW; 517F9B7F6B960A6A CRC64;

Query Match 100.0%; Score 897; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 1.5e-52;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKPEFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 60
Db 1912 KIVVKDFARNTTVKPEFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 1971
QY 61 ELEKGYPFGWISGPEGKDGAGYVNLKDTFKVPFKKIEEKKKEENKPTFDVSKKCD 120
Db 1972 ELEKGYPFGWISGPEGKDGAGYVNLKDTFKVPFKKIEEKKKEENKPTFDVSKKCD 2031
QY 121 NPQVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 174
Db 2032 NPQVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2085

RESULT 2
Q97RY6 STREP PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE Serine protease, subtilase family.
GN OrderedLocusNames=SP0641;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzaple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Lofus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AS007373; AAK74791.1; -; Genomic_DNA.
DR PIR; F95074; F95074.
DR HSSP; P00782; 2SBT.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0042809; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.

```

```

DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS0847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

Query Match 100.0%; Score 897; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.5e-52;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVKDFARNTTVKPEFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 60
Db 1933 KIVVKDFARNTTVKPEFILNKDTGVSSELKPHRVTVTTIQNGKMSSTIVSEEDFILPVYKG 1992
QY 61 ELEKGYPFGWISGPEGKDGAGYVNLKDTFKVPFKKIEEKKKEENKPTFDVSKKCD 120
Db 1993 ELEKGYPFGWISGPEGKDGAGYVNLKDTFKVPFKKIEEKKKEENKPTFDVSKKCD 2052
QY 121 NPQVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 174
Db 2053 NPQVNHSQLNESHKREDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2106

RESULT 3
Q8DQP7 STRR6 PRELIMINARY; PRT; 2144 AA.
AC Q8DQP7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN Name=prtA; OrderedLocusNames=spr0561;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Eszrem S.T., Fritz L., Fu D.-J., Fuller W., Geisinger C.,
RA Gilmore R., Glase J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McLaren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008434; AAK9365.1; -; Genomic_DNA.
DR PIR; A97942; A97942.
DR HSSP; P00782; 2SBT.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.

```

DR	PRINTS; PR00723; SUBTILISIN.
DR	TIGRFAMS; TGR01167; LPXTG anchor; 1.
DR	PROSITE; PSS0847; GRAM POS ANCHORING; 1.
DR	PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR	PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR	PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW	Cell wall; Signal.
FT	SIGNAL 1 Potential.
FT	CHAIN 20 2144 cell wall-associated serine proteinase PrtA.
SQ	SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;
Query Match	99.6%; Score 893; DB 2; Length 2144;
Best Local Similarity	98.9%; Pred. No. 2.8e-52;
Matches 172; Conservative 2; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 KIVVKDFARNTTVTKETPLANKOTCEVSELKPHRVVTVIQNGKMSSTIVSEEDFILPVYKG 60       
Dd	1937 KIVVKDFARNTTVTKETPLNKOTCEVSELKPHRITVTIQTQNGKMSSTIVSEEDFILPVYKG 1996       
Qy	61 ELEKGYPQDGEWISGPEGKKDAGVYNLSKDTFIKPVPFKKIEBKKEENKPTFDVSKKD 120       
Dd	1997 ELEKGYPQDGEWISGPEGKKDAGVYNLSKDTFIKPVPFKKIEBKKEENKPTFDVSKKD 2056       
Qy	121 NPQVNSQLNESHKREDLQREHSQKSDSTKVQTTATVLDKNNTSKSSTTNPNK 174       
Dd	2057 NPQVNSQLNESHKREDLOREDSQKSDSTKVQTTATVLDKNNTSKSSTTNPNK 2110       

RESULT 5

Q4XUI6 PLACH PRELIMINARY; PRT; 300 AA.

Q4XUI6

AC Q4XUI6

DT 13-SEP-2005 (TrEMBLrel. 31, Created)

DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Hypothetical protein (Fragment).

GN ORFNames=PC000286.03;

OS Plasmodium chabaudi.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCEI\_TaxID=5825;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,

RA Berriaman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,

RA James K., Rutherford K., Harris B., Harris D., Churcher C.,

RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,

RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,

RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;

RT "A comprehensive survey of the plasmodium life cycle by genomic,

RT transcriptomic, and proteomic analyses.";

RL Science 307:82-86(2005).

CC -1- CAUTION: The sequence shown here is derived from an

CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is

CC preliminary data.

CC ENBL; CAAJ01003049; CAH79425.1; -; Genomic\_DNA.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 4.

DR PRINTS; PR00320; GPROTEINRPT.

DR ProDom; PD000018; WD40; 3.

DR SMART; SM00320; WD40; 5.

DR PROSITE; PS00678; WD\_REPEATS\_1; 1.

DR PROSITE; PS00082; WD\_REPEATS\_2; 4.

DR PROSITE; PS02034; WD\_REPEATS\_REGION; 1.

KW Hypothetical protein; Repeat; WD repeat.

FT NON\_TER 1

FT 1

SQ SEQUENCE 300 AA; 34469 MW; 8EDR4E512AFB1945 CRC64;

Query Match 15.5%; Score 139; DB 2; Length 300;

Best Local Similarity 23.2%; Pred. No. 0.11;

Matches 45; Conservative 38; Mismatches 61; Indels 50; Gaps 8;

QV 1 KIVVDQFARNTTVKKEPILNKDTGE-----VSELKPHRVTVTIQNGKEM 43

```
Db      85 KTIYNNFLKNQNHKLVSGDDGTLHLIELCKNDKYKSTRLLGHQKPVHTIQSPNGKFI 144
Qy      44 STIYSE-----EDFILPVYKGELEKGYQFDGWEISGPEKQDAGVIVNLSKDTIK 95
Db      145 ASSSFDKSIIRISGIDGTGYLAVYRGHVGPAYKI-AMSI-----DNNYIIVSCSDSTLK 196
Qy      96 -----PVFKKIEEKE---EENKPTFVDSKKDNPQVNHSQLNESHREKEDLQREHS 144
Db      197 LWRINHLVPLLRKKEENAEQTDEQK-----NEQENPQ-VNDQPDANSEKKKKNEK 250
Qy      145 QKSDSTKQVATVL 158
Db      251 ERNDKTKNKIKTL 264
```

## RESULT 6

```
Q8CPK8 STABP
ID Q8CPK8 STABP PRELIMINARY; PRT; 775 AA.
AC Q8CPK8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2004 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Penicillin-binding protein 1.
GN OrderedLocusNames=SE0856;
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 12228;
RX PubMed=12950922; DOI=10.1046/j.1365-2958.2003.03671.x;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593 (2003).
DR EMBL; AE016746; RAO04453.1; -; Genomic_DNA.
DR HSP; F14677; IQME.
DR GO; GO:0008658; F:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR005311; PBP dimer.
DR InterPro; IPR012338; PBP_trypt_fold.
DR InterPro; IPR001460; Pencil_bind_tpept.
DR Pfam; PF03793; PASTA; 2.
DR Pfam; PF00905; Transpeptidase; 1.
DR SMART; SM00740; PASTA; 2.
KW Complete proteome.
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;
```

```
Query Match 13.3%; Score 119; DB 2; Length 775;
Best Local Similarity 27.0%; Pred. No. 7.2;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;

Qy 20 KDTGEVSELKPHRVTVTIQNGKEMSGSTIVSEDFILPVYK-----GLEKGYQFDGW-- 71
Db 614 EDSVNAQSLKP---ITIGNGQIKQSVKSGTKVLPKSHKVMMLTMDGLTNP-DMTGWTK 668
Qy 72 -HISGPE-----GKGDAGYIN--LSKDTIKPVFKKIEEKEENKPTFVDS----K 117
Db 669 EDVLAFEDTKIKVSTKGNGFTVNOSISKGQIHK-----NKKIEVLSLSD 715
Qy 118 KKDNPQVNHSQLNESHREKEDLQREHSQKSDSTKQVATVLKNNISSKSTTN 171
Db 716 TDDQEKTDSDSNKSKDKADEHSNTSSSTKN-----DKSNADSKNDSDD 763
```

## RESULT 7

```
Q512T7 ENTHI
ID Q512T7 ENTHI PRELIMINARY; PRT; 1069 AA.
```

```
AC Q512T7;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Fimbriae-associated protein, putative.
GN ORFNames=89.t00007;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Amedeo P., Rongaglia P., Berriman M., Hirt R.P., Mann B.J., Noraki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Church C., Hance Z., Harris B., Squares R., Whitehead S.,
RA Jagels K., Moule S., Mungall K., Ormond D., Norbertzak H., Price C., Wang Z.,
RA Quail M.A., Rabinowitsch E., Norbertzak H., Norbertzak H., Lohia A.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sichertitz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868 (2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFB01000328; EAL47849.1; -; Genomic DNA.
SQ SEQUENCE 1069 AA; 120249 MW; 1D38E4F0A8759CB4 CRC64;

Query Match 13.0%; Score 117; DB 2; Length 1069;
Best Local Similarity 25.6%; Pred. No. 14;
Matches 50; Conservative 39; Mismatches 70; Indels 36; Gaps 9;

Qy 1 KIVVKDFARNTT-----VKEFILNKDT-GEVSELKPHRVTVTIQNG-KEMSGSTIVSEED 52
Db 108 EIEQKEIEKNTTSTIEDESVKSLDSDSKQSLPEVSKDNVNESSSTLTNGEKKLSLCLNEQD 167
Qy 53 FILPVYKGELEKGYQFDGWEISGPEKQDAGVIVNLSKDTIKPVFKKIEEKEENK-- 110
Db 168 -----ELQSKSSSTD--NKNKRDRIHFVDVLPKNEKEISMEIESKTEEEKSN 217
Qy 111 ---PTFDVSKKKNPQ-----VNHSQLNESHREKEDLQREHS-QKSDSTKD-----VT 154
Db 218 LQIFSLNLSEKQKKNVSEIATVLLKSNSSNGEEDKQDEEVSCERKPFDSQEEKKEEMIK 277
Qy 155 ATVLDDKNNISSKSTT 169
Db 278 AEVSQNKVEKDKSTT 292

RESULT 8
Q50LX8 ENTHI
ID Q50LX8 ENTHI PRELIMINARY; PRT; 296 AA.
AC Q50LX8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=657.t00001;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Amedeo P., Rongaglia P., Berriman M., Hirt R.P., Mann B.J., Noraki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
```



```

DE Merozoite surface protein 3 (Fragment).
GN Name=msp3;
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;
RA Okeru D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252286; CAB65754.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
FT NON_TER 1 346
FT SEQUENCE 346 AA; 39127 MW; A804B96BDFAFA010 CRC64;
SQ
Query Match 12.5%; Score 112.5; DB 2; Length 346;
Best Local Similarity 23.7%; Pred. No. 8.4;
Matches 42; Conservative 33; Mismatches 67; Indels 35; Gaps 7;

QY 10 NTTVKEFILN-KDTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGQYF 68
D 156 STKTKEAQAQNAKAYQKNAQAVLKAKKASS-----YNYL----- 197
QY 69 DGWEISGF--EGKDGAG-----YVINLSKDTFIKPVFKKIEEKEE-----E 108
D 195 -GWEEGGVPEHKKENMLSHLYVSSKDKENISKENDVDLDEKEEAEETEELKEEKN 253
QY 109 NKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNISSKST 168
D 254 EETESEINDEBEKEEENDEKKEQKQESQNNNDQKQKQMEAE-----QNLISKQ 308
QY 169 TNN 171
D 309 NNN 311
SQ
RESULT 13
Q81J55_PLAF7 PRELIMINARY; PRT; 354 AA.
AC Q81J55;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2004 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Merozoite surface protein 3.
GN ORFNames=PF10_0345;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlton J.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S., Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B., Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium falciparum".
RL Nature 419:458-511(2002).
DR EMBL; AB014834; AAN35542.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 12.5%; Score 112.5; DB 2; Length 354;
Best Local Similarity 22.4%; Pred. No. 8.6;
Matches 41; Conservative 35; Mismatches 62; Indels 45; Gaps 7;

QY 10 NTTVKEFILN-KDTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGQYF 68
D 153 STKTKEAQAQNAKAYQKNAQAVLKAKKASS-----YNYL----- 194
QY 69 DGWEISGF--EGKDGAG-----YVINLSKDTFIKPVFKKIEEKEE-----E 108
D 195 -GWEEGGVPEHKKENMLSHLYVSSKDKENISKENDVDLDEKEEAEETEELKEEKN 253
QY 109 NKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNISSKST 168
D 254 EETESEINDEBEKEEENDEKKEQKQESQNNNDQKQKQMEAE-----QNLISKQ 308
QY 169 TNN 171
D 309 NNN 311
SQ
RESULT 14
Q81J55_PLAF7 PRELIMINARY; PRT; 354 AA.
AC Q81J55;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2004 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Merozoite surface protein 3.
GN ORFNames=PF10_0345;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., Carlton J.M., Pain A., Nelson K.B., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S., Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S., Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B., Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium falciparum".
RL Nature 419:458-511(2002).
DR EMBL; AB014834; AAN35542.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Merozoite.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 12.5%; Score 112.5; DB 2; Length 354;
Best Local Similarity 22.4%; Pred. No. 8.6;
Matches 41; Conservative 35; Mismatches 62; Indels 45; Gaps 7;

QY 10 NTTVKEFILN-KDTGEVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEKGQYF 68
D 153 STKTKEAQAQNAKAYQKNAQAVLKAKKASS-----YNYL----- 194
QY 69 DGWEISGF--EGKDGAG-----YVINLSKDTFIKPVFKKIEEKEE-----E 108
D 195 -GWEEGGVPEHKKENMLSHLYVSSKDKENISKENDVDLDEKEEAEETEELKEEKN 253
QY 109 NKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNISSKST 168
D 254 EETESEINDEBEKEEENDEKKEQKQESQNNNDQKQKQMEAE-----QNLISKQ 308
QY 169 TNN 171
D 309 NNN 311
SQ

```





```
Db 608 K-----KKCKKNDKIKSIKIKTL 626

RESULT 17
Q815F3 PLAF7 PRELIMINARY; PRT; 829 AA.
AC Q815F3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PFL11275C;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]

RX MEDLINE=22255705; PubMed=1236864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perlea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
RN [2]

RP NUCLEOTIDE SEQUENCE.
RA Hyman R.W., Fung E., Conway A., Kurd O., Mao J., Miranda M.,
RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AS014848; AAN36341.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 829 AA; 98815 MW; BF2675E301B2CE93 CRC64;

Query Match 12.5%; Score 112; DB 2; Length 829;
Best Local Similarity 24.0%; Pred. No. 23;
Matches 44; Conservative 39; Mismatches 58; Indels 42; Gaps 9;

QY 8 ARNTVKFELKNDTGEVSELKPHRVTVTIQKEMSSITIVSEEDFILPVYKGELEKGVQ 67
Db 487 AATITSNDSILKNTNKKI--IQP--LEYLLKNGD--KTLMTERDIVLDVFPHPYMKKXH 540
QY 68 FDGWEISGPEGKDGAVINLSKDTPIKVPFKKIEKKEENKPT-----F 113
Db 541 LN-----KKETLFPNLSN-----PREIEKKNKDKKGTTHNNKNDAEYMLKY 583
QY 114 DVSKKKDN-PQNVHSQLNESH-RKEDLQREHSQKSDSTKVTVATVLDKNNISKSSTNN 171
Db 584 KIKKKKKNFEENNTELNDNSIKKNNKLVEH---DNSLKQRIIINDKNVIEHTKIYDN 640
QY 172 PNK 174
Db 641 QKK 643

RESULT 18
Q25706 PLAF6 PRELIMINARY; PRT; 379 AA.
AC Q25706;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polymorphic antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.

Query Match 12.4%; Score 111.5; DB 2; Length 775;
Best Local Similarity 24.9%; Pred. No. 23;
Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;

QY 20 KDTGEVSELPKPHRVTVTIQKEMSSITIVSEEDFILPVYK-----GBLEKGYQPDGW-- 71
Db 171 KPSRINLFSRKTKYAEQV--EKDYERAKNAYQKQANQAVLKAKKASSYDYILGWEFGGV 228
QY 77 -EGKKDAG-----YVINLSKDTPIKVPFKKIEKKEENKPTFDVSKKDNPOVNHSQLN 130
Db 229 PEHKKENMLSHLVSSKDKENISKENDVDLDE-KEEAEETEEBELEKEEETESIS 287
QY 131 -----ESHRKEDLQREHSQKSDSTKVTVATVLDKNNISKSSTNN 171
Db 288 EDEBESEEEKEEENDKKKEQKESNNENNDDKKQMEA-----QNLISKNNQNNN 336

RESULT 19
Q5HQ11 STAEQ PRELIMINARY; PRT; 775 AA.
AC Q5HQ11;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Penicillin-binding protein 1.
GN Name=pbp1; OrderedLocusNames=SERP0746;
OS Staphylococcus epidermidis (strain ATCC 35984 / RP62A).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=116279;
RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=1574986; DOI=10.1128/JB.187.7.2426-2438.2005;
RA Gill S.R., Fouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T.,
RA Ravel J., Paulsen I.T., Kolonay J.F., Brinkac L.M., Beanan M.J.,
RA Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S.,
RA Hatt D.H., Vamathevan J.J., Khouri H., Utterback T.R., Lee C.,
RA Dimitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H.,
RA Hance I.R., Nelson K.E., Fraser C.M.;
RT "Insights on evolution of virulence and resistance from the complete
RT genome analysis of an early methicillin-resistant Staphylococcus
RT aureus strain and a biofilm-producing methicillin-resistant
RT Staphylococcus epidermidis strain."
RL J. Bacteriol. 187:2426-2438(2005).
DR EMBL; CP000029; AAW54126.1; -; Genomic_DNA.
DR TIGR; SERP0746; -;
DR GO; GO:0008658; P:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR005311; PBP dimer.
DR Pfam; PF03793; PASTA; 2.
DR Pfam; PF03717; PBP dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR SMART; SM00740; PASTA; 2.
KW Complete proteome.
SQ SEQUENCE 775 AA; 86352 MW; B9395893E0043694 CRC64;

Query Match 12.4%; Score 111.5; DB 2; Length 775;
Best Local Similarity 24.9%; Pred. No. 23;
Matches 42; Conservative 28; Mismatches 66; Indels 33; Gaps 8;

QY 20 KDTGEVSELPKPHRVTVTIQKEMSSITIVSEEDFILPVYK-----GBLEKGYQPDGW-- 71
Db 171 KPSRINLFSRKTKYAEQV--EKDYERAKNAYQKQANQAVLKAKKASSYDYILGWEFGGV 228
QY 77 -EGKKDAG-----YVINLSKDTPIKVPFKKIEKKEENKPTFDVSKKDNPOVNHSQLN 130
Db 229 PEHKKENMLSHLVSSKDKENISKENDVDLDE-KEEAEETEEBELEKEEETESIS 287
QY 131 -----ESHRKEDLQREHSQKSDSTKVTVATVLDKNNISKSSTNN 171
Db 288 EDEBESEEEKEEENDKKKEQKESNNENNDDKKQMEA-----QNLISKNNQNNN 336
```



Query Match	12.3%	Score 110;	DB 2;	Length 500;
Best Local Similarity	22.4%;	Pred. No. 19;		
Matches	43; Conservative	34; Mismatches	75; Indels	40; Gaps
6;				
14	KPFLNKDTGVESELKPHRVTTVIONGKEMSS-----TIVSEEDFILPVVK	60		
	:   :	:   :	:	:   :
b	11 KQDLLKKAKETRE-EPFTYYPINVAESEKKLKKVPWNEHTATSDDKILIDATLG	69		
	:   :	:	:	:   :
61	ELEKGVPDGVISGFEGKKDAGYVN-----LSKDTFIKVPFK-----KIEKK	105		
y	:   :	:	:	:   :
b	70 NEEAREVQIYQITFANVORSTAYLTQEGVPQRDPDFEVEMFKSPQMCKINKLIEKR	129		
	:   :	:	:	:   :
106	EENKPTEFDVS--KCONPNHSQLNESHRK-----EDLREEHSQKSOSTKDVT	154		
y	:     :	:	:	:     :
b	130 DENKKKEEKKSHSKLKNQMKIKKWEERHEKKQNVAIBQWKCEIKEGSGKARDLD	189		
	:     :	:	:	:     :
155	ATVLDRKNNTSSK	166		
y	:			
b	190 EIIKKONNOISKK	201		

RESULT	23
Q7RLE7	PLAYO PRELIMINARY; PRT; 674 AA.
C	Q7RLE7; 7RELE7 PLAYO
T	01-WAR-2004 (TrEMBLrel_26, Created)
T	01-WAR-2004 (TrEMBLrel_26, Last sequence update)
T	01-WAR-2004 (TrEMBLrel_26, Last annotation update)
E	Notchless-related.
N	Name=PY02598;
S	Plasmodium yoelii yoelii.
X	Kyokyoita; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
N	[1]_TaxID=73239;
P	NUCLEOTIDE SEQUENCE.
C	STRAIN=17XNL;
X	MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
A	Carlton J.M., Angluoli S.V., Suh B.B., Koij T.W., Pertea M.,
A	Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
A	Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Bidwell S.L.,
A	Shallom S.J., van Aken S.B., Riedmuller S.B., Feldlyum T.V.,
A	Cho J.K., Quackenbush J., Sedegah M., Shoaihi A., Cummings L.M.,
A	Florns L., Yates J.R. III, Raine J.D., Sindien R.E., Harris M.A.,
A	Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
A	van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
A	Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
A	Carucci D.J.;
T	"Genome sequence and comparative analysis of the model rodent malaria
T	parasite Plasmodium yoelii yoelii.";
L	Nature 419:512-519(2002).
C	-!- CAUTION: The sequence shown here is derived from an
C	EMBL/GenBank/DDJB whole genome shotgun (WGS) entry which is
C	preliminary data
R	EMBL; ABAL01000712; EAA22065.1; -; Genomic_DNA.
R	HSSP; Pf6649; IERJ.
R	InterPro; IPR001680; WD40.
R	Pfam; PF00400; WD40; 8.
R	PRINTS; PR00320; GPROTEINBRPT.
R	ProDom; PD000018; WD40; 4.
R	PROSITE; PS00678; WD REPEATS 1; 2.
R	PROSITE; PS50082; WD_REPEATS_2; 6.
R	FORSITE; PS50294; WD_REPEATS_REGION; 2.
R	Repeat; WD repeat.
Q	SEQUENCE 674 AA; 76535 MW; E7521B469FE8E0F7 CRC64;

```

Query Match      12.3%; Score 110; DB 2; Length 674;
Best Local Similarity 20.8%; Pred. No. 26;
Matches 42; Conservative 34; Mismatches 74; Indels 52; Gaps 7
y 1 KIWVKDPARNTVTKPEILNKDTGE-----VSELKPHRVTVIQNGEM 43
b 445 KNIYNNFLKNQHEKLVSGSDGTGLHLIECLXNDKYKTRLLGHOKPVITHQFSPNGKFI 504

```

Qy	44	SSTIVS-----EDFILPVYKGELEKGYQDFGWEISGFEKGKDAYVINLSKDTFIK	95
	:	:	:
Db	505	ASSFDSKSIIRWSGIDGTYLAVPRGHVPYKI-AWSI-----DNNYIISCSQDSTLK	556
Qy	96	-----PVPKTEE-----KKEENKPTFDV--SKKDNQPNVHNSQLNESHKKE	136
	:	:	:
Db	557	LWRINHVLVPLLLKKKEBNGDQPKSPDKNDQKDKDQKDDQKDKNDNDHANNQEDGE	616
Qy	137	DLOREHSQKSDSTKDVATVTL	158
	:	:	:
Db	617	EKKKKKEKKDKIKSIKILL	638
	:	:	:
RESULT 24			
Q50VJ0	ENTH		
ID	Q50VJ0_ENTH1	PRELIMINARY;	PRT; 384 AA.
AC	Q50VJ0;		
DT	13-SEP-2005	(TrEMBLrel. 31, Created)	
DT	13-SEP-2005	(TrEMBLrel. 31, Last sequence update)	
DT	13-SEP-2005	(TrEMBLrel. 31, Last annotation update)	
DE	HMGB box protein.		
GN	ORFName=188 t00012;		
OS	Entamoeba histolytica HM-1:IMSS.		
OC	Eukaryota; Entamoebidae; Entamoeba.		
OX	NCBI_TaxId=294381;		
FN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=HM-1:IMSS;		
RX	PubMed=15729342; DOI=10.1038/nature03291;		
RA	Lotus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,		
RA	Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,		
RA	Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,		
RA	Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,		
RA	Chillingworth T., Church C., Hance Z., Harris B., Harris D.,		
RA	Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,		
RA	Quail M.A., Rabinowitz E., Norbertczak H., Price C., Wang Z.,		
RA	Gullien N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,		
RA	Poster P.G., Sichevits-Ponten T., Weber C., Singh U., Mukherjee C.,		
RA	El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,		
RA	Fraser C.M., Hall N.;		
RT	"The genome of the protist parasite Entamoeba histolytica.";		
RL	Nature 433:865-868(2005).		
CC	!- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; AAFB01000585; EAL45607.1; -; Genomic DNA.		
SQ	SEQUENCE 384 AA; 45464 MW; 207789F65D72B019 CRC64;		
Query Match 12.2%; Score 109.5; DB 2; Length 384;			
Best Local Similarity 27.1%; Pred.No.15;			
Matches 39; Conservative 28; Mismatches 64; Indels 13; Gaps 4			
Qy	27	ELKPHRYTVVTIQNGKEMSSITVSEDFILPVYKGELEKGYQFD--GWEISGFEKGKDAYG	84
	:	:	:
Db	226	EETKVKVEIKKEDDEKTKVVEIKED-----EKKEKHSKKEDKKKEMKKNGKKS	280
Qy	85	VINLSKDTFIKPVPKTEEKKEENKPTFDVSKKDNQPNVHNSQLNESHKREDLQREHS	144
	:	:	:
Db	281	KEDTKDKKVKVSKSEKSDIKKEDEKH----EKKEENTEEKKPKPESEKESKKEKKS	336
Qy	145	QKSDSTKD--VTATVLDKNNISSK	166
	:	:	:
Db	337	KKEDKKDEKSKKVEDKSKKKQK	360

RESULT 25	
Q8I436 PLAF7	
ID Q8I436 PLAF7 PRELIMINARY;	PRT; 3008 AA.
AC Q8I4361	
DT 01-MAR-2003 (TRENBLrel. 23, Created)	
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)	
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)	
DE Hypothetical protein PFE0325w.	







```
DR PRINTS; PRO1182; ORNDCRBLXASE.
SQ SEQUENCE 948 AA; 110350 MW; 43F103DB83F12835 CRC64;

Query Match 12.0%; Score 107.5; DB 2; Length 948;
Best Local Similarity 22.0%; Pred. No. 55;
Matches 54; Conservative 32; Mismatches 71; Indels 89; Gaps 11;

QY 4 VKDFARNTTVKFFLNKDTGVESEL-----KPHRVTVTIQNGKMSSTIVSEEDPI 54
DB 547 VEDMSSNWGFNFIYINLGGYFEELEYDNAKHDKLHYCTLQLQEIKKDIQIFLNEETPL 606
QY 55 LPVY-----KGELEKGYQFDGWEISGFEKGD-----AGYVINSKDTFIKPVFKK 100
DB 607 KTKYGYVSFEKISLAINMSIDY----FSHMKDNLRVICPEGRYVVAASSTLAVKIGR 662
QY 101 IEKKEEENKPTF-----DVSKKDNQPNVHSQLNESHKED----- 137
DB 663 -----RPTFQGIIMLKDLKAHYDPLNFAQENKKQDEPKINNNNDNNNDNNN 713
QY 138 LQREHSQK-----SDST-----KDVTVTLVDK--NNIS-SKS 167
DB 714 NNNNNNQGGGNNNDLIITSTNDSTNNKNDHSSQVIQNVCTIRDKGDNKINTH 773
QY 168 TTNPN 173
DB 774 TINPN 779

RESULT 33
Q640L5 MOUSE
ID Q640L5 MOUSE PRELIMINARY; PRT; 1455 AA.
AC Q640L5;
DT 25-OCT-2004 (Tremblrel. 28, Created)
DT 25-OCT-2004 (Tremblrel. 28, Last sequence update)
DE Sarcoma antigen NY-SAR-41.
GN Names=Ccd18; Synonyms=4932411G06Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Head;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grumwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Head;
RA Director MGC Project;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC082598; AAH82598.1; -; mRNA.
DR Ensembl; ENSMUSG00000056531; Mus musculus.

DR MGI; MGI:1922974; 4932411G06Rik.
DR MGI; MGI:1922974; Ccdc18.
DR GO; GO:0009279; C:outer membrane (sensu Gram-negative Bacteria); IEA.
DR GO; GO:0006952; P:defense response; IEA.
DR InterPro; IPR001800; Lipoprotein_6.
SQ SEQUENCE 1455 AA; 169741 MW; EF04ACB9E4AA2472 CRC64;

Query Match 12.0%; Score 107.5; DB 2; Length 1455;
Best Local Similarity 24.5%; Pred. No. 87;
Matches 39; Conservative 31; Mismatches 50; Indels 39; Gaps 5;

QY 14 KEFINKDTGVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGBLEKGYQPDGWEI 73
DB 1084 KEFIMLQNEISQLK-KEIERTQRMKEMESVKEQEDYATQYKEVI----- 1131
QY 74 SGFEKGDAGVIVINSKDTFTKPVFKIETKEE-----ENKPTFDVSKKD- 120
DB 1132 -----DLGQLRLTQEQMONTSELVEARRQEQVQAQREIERLAGELEDIKLSKEKA 1184
QY 121 -----NPQVHSQLNESHKEDLQRE--EHSQKSDSTKD 152
DB 1185 HGNRLABELGASQVRAHLEARMQAEIKGLSSEVDSLKE 1223

RESULT 34
QSV9M0 PLAKN
ID QSV9M0 PLAKN PRELIMINARY; PRT; 374 AA.
AC QSV9M0;
DT 01-FEB-2005 (Tremblrel. 29, Created)
DT 01-FEB-2005 (Tremblrel. 29, Last sequence update)
DT 01-FEB-2005 (Tremblrel. 29, Last annotation update)
DE Merozoite surface protein 5.
GN Name=MSP5;
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=H;
RA Black C.G., Wang L., Topolaka A.E., Finkelstein D.I., Horne M.K.,
RA Thomas A.W., Mohandas N., Coppel R.L.;
RT "Merozoite surface proteins 4 and 5 of Plasmodium knowlesi have
differing cellular localisation and association with lipid rafts."
RL Mol. Biochem. Parasitol. 138:153-158 (2004).
RE EMBL; AY573058; AAT7929.1; -; Genomic_DNA.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Merozoite.
SQ SEQUENCE 374 AA; 41642 MW; COD687C6F23FE989 CRC64;

Query Match 11.9%; Score 106.5; DB 2; Length 374;
Best Local Similarity 25.8%; Pred. No. 23;
Matches 39; Conservative 28; Mismatches 63; Indels 21; Gaps 6;

QY 38 QNGKMSSTIVSEEDFILPVYKGEI--EKG-----YQPDGWEISGFEKGDAGVIVINL-- 88
DB 19 QSGHPLKFSFWREKIHLOIYTNLLREKNDVQEMESPSISGTEGKQIQLSHLQL 78
QY 89 -----SKDTFTKPVFKIETKEEENKPTFDVSKKDNQPNVHSQLNESHKEDLQREH 143
DB 79 QSGKHQVSVFLSANDSNLKGANEASASBENGKKSDEENVCKSD--EENAKKSD---BEN 134
QY 144 SQKSDSTKDVTVTLVDKNNISKS---TTNN 171
DB 135 KDANSTKDAESABGEENPVSQENQMKTLNN 165

RESULT 35
QSW5T1 TETPY
ID QSW5T1 TETPY PRELIMINARY; PRT; 1015 AA.
AC QSW5T1;
DT 01-FEB-2005 (Tremblrel. 29, Created)
DT 01-FEB-2005 (Tremblrel. 29, Last sequence update)
```









**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 102.913 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848

Sequence: 1 TVVKEFILKDTGVSELKP.....ATVLDKNMISSKSTNNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	848	100.0	773	4	AAB48343	Aab48343 S. pneumo
2	848	100.0	2120	3	AAY81710	Aay81710 Streptoco
3	848	100.0	2140	6	ABU01020	Abu01020 S. pneumo
4	848	100.0	2140	6	ABU45746	Abu45746 Protein e
5	848	100.0	2140	8	ADM92113	Adm92113 S. pneumon
6	848	100.0	2140	8	ADT50099	Adt50099 S. pneumon
7	845	99.6	637	8	Adt-94534	Novel S.
8	845	99.6	637	9	AEA58404	Aea58404 Streptoco
9	845	99.6	2138	8	ADK48759	Adk48759 Streptoco
10	615	72.5	117	2	AAW55096	Aaw55096 Streptoco
11	615	72.5	117	5	ABP54590	Abp54590 S. pneumo
12	615	72.5	117	7	ADC45149	Adc45149 S. pneumo
13	119	14.0	746	4	AAg81779	AAg81779 S. epide
14	119	14.0	778	5	ABF39023	ABf39023 Staphyloc
15	119	14.0	778	8	ADS06368	AdS06368 Staphyloc
16	111.5	13.1	354	9	ADT27253	Adt27253 Plasmodiu
17	111.5	13.1	775	6	ABU42797	Abu42797 Protein e
18	111	13.1	707	6	ABU25018	Abu25018 Protein e
19	110.5	13.0	647	9	ADT279635	Adt279635 P. falcip
20	110.5	13.0	651	8	AD019012	Ado19012 Amino aci
21	110.5	13.0	651	8	AD019010	Ado19010 P. falcip
22	108	12.7	188	9	ADT279639	Adt279639 P. falcip
23	107.5	12.7	470	8	ADT56185	Adt56185 Plant pol
24	107.5	12.7	484	3	AAG47777	Arabidops

25	106	12.5	665	3	AAB18278	Aab18278 Plasmodiu
26	106	12.5	665	7	ABO23606	Abo23606 Plasmodiu
27	103.5	12.2	169	9	ADZ79634	Adz79634 P. falcip
28	103	12.1	903	6	ABU24404	Abu24404 Protein e
29	101.5	12.0	564	4	ABB61977	Abb61977 Drosophil
30	100	11.8	1791	8	ADP25441	Adp25441 Plasmodiu
31	99.5	11.7	1384	6	ABP55413	Abp55413 Human MDD
32	99.5	11.7	1404	6	ABP55393	Abp55393 Human MDD
33	97.5	11.5	645	9	ADW88460	Adw88460 Staphyloc
34	97.5	11.5	645	9	ADW88459	Adw88459 Staphyloc
35	97.5	11.5	645	9	ADW88458	Adw88458 Staphyloc
36	97.5	11.5	1184	6	ABU25330	Abu25330 Protein e
37	97.5	11.5	2468	6	ABR64281	ABr64281 Angiogene
38	97.5	11.5	2468	7	ADE62723	AdE62723 Human Pro
39	97.5	11.5	2468	7	ADE62719	AdE62719 Human Pro
40	97.5	11.5	2468	7	ADE62727	AdE62727 Human Pro
41	97.5	11.5	2468	7	ADE62715	AdE62715 Human Pro
42	97.5	11.5	2468	8	ADL12997	Adl12997 Human ste
43	97.5	11.5	2468	8	ADN05260	Adn05260 Antipsori
44	97.5	11.5	2468	8	ADR14614	Adr14614 Human NF-
45	97.5	11.5	2519	4	ABG16636	Abg16636 Novel hum

#### ALIGNMENTS

##### RESULT 1

ID	AAB48343	standard; protein; 773 AA.
XX	AAB48343;	
AC	AAB48343;	
DT	20-APR-2001	(first entry)
XX		
DE	S. pneumoniae	Spl130 polypeptide.
XX		
KW	Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;	
KW	bronchial; lung; blood; infection; immune response; immunotherapy;	
KW	antibacterial; auditory; vaccine.	
XX		
OS	Streptococcus pneumoniae.	
XX		
FN	WO200076540-A2.	
XX		
PD	21-DEC-2000.	
XX		
PF	09-JUN-2000; 2000WO-US015925.	
XX		
PR	10-JUN-1999; 99US-0138453P.	
XX		
PA	(MEDI-) MED IMMUNE INC.	
XX		
PI	Adamou JE, Choi GH;	
XX		
DR	WPI; 2001-112197/12.	
XX		
DR	N-PSDB; AAC84742.	
XX		
PT	New vaccines comprising Spi28 or Spi30 polypeptides, for treating and	
PT	preventing pneumococcal infections, particularly infections caused by	
PT	Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or	
XX	blood infections.	
XX		
PS	Claim 8; Page 51-54; 54pp; English.	
XX		
CC	The invention relates to novel immunogenic polypeptides, Spi28 and Spi30	
CC	from S. pneumoniae. Vaccines comprising the polypeptides are useful for	
CC	the treatment and prevention of pneumococcal infections, particularly	
CC	infections caused by Streptococcus, such as otitis media, nasopharyngeal,	
CC	bronchial, lung or blood infections. The antigens are used as immunogenic	
CC	agents to stimulate an immune response. The antisera and antibodies may	
CC	also be used in diagnosing and treating pneumococcal infections.	
CC	Recombinant polypeptides serve as a mechanism for stimulating production	
CC	of antibodies for use in passive immunotherapy, diagnostic reagents, and	

CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Spl30 polypeptide  
XX  
SQ Sequence 773 AA;

Query Match 100.0%; Score 848; DB 4; Length 773;  
Best Local Similarity 100.0%; Pred. No. 2.8e-74;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKGFILNKDGTGEVSELKPHRVTTVITQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 60  
Db 610 TTVKGFILNKDGTGEVSELKPHRVTTVITQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 669  
QY 61 WISGFEKGDAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKKNQPNVHNSQLN 120  
Db 670 WISGFEKGDAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKKNQPNVHNSQLN 729  
QY 121 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 164  
Db 730 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 773

RESULT 2  
AAY81710  
ID AAY81710 standard; protein; 2120 AA.  
XX  
AC AAY81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW Bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PF 27-JUL-1999; 99WO-GB002452.  
XX  
XX 27-JUL-1998; 98GB-00016336.  
PR 19-MAR-1999; 99US-0125329F.  
XX  
XX (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
PI Le Page RWP, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
XX WPI; 2000-195301/17.  
DR N-PSDB; AA291806.  
XX  
PT Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections.  
XX  
PS Claim 2; Page 41-42; 76pp; English.  
XX  
XX This sequence represents a Streptococcus pneumoniae protein of the  
XX invention. The proteins (or their homologues, derivatives and/or  
XX fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
XX compositions comprising the proteins are useful as vaccines and also in  
XX diagnostic assays. The sequences are useful for the detection or  
XX diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
XX with them. Agents capable of antagonising, inhibiting or interfering with  
XX the function or expression of the protein or polypeptide are useful in  
XX medical compositions in the treatment or prophylaxis of S. pneumoniae  
XX infection. As the sequences can be used to treat S. pneumoniae infection,  
XX they can be used to treat bacterial pneumonia, which has high rates in  
XX young children, the elderly, and in patients with predisposing conditions  
XX such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis  
XX  
SQ Sequence 2120 AA;

Query Match 100.0%; Score 848; DB 3; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 1.1e-73;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKGFILNKDGTGEVSELKPHRVTTVITQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 60  
Db 1923 TTVKGFILNKDGTGEVSELKPHRVTTVITQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 1982  
QY 61 WISGFEKGDAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKKNQPNVHNSQLN 120  
Db 1983 WISGFEKGDAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKKNQPNVHNSQLN 2042  
QY 121 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 164  
Db 2043 ESHRKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 2086

RESULT 3  
ABU01020  
ID ABU01020 standard; protein; 2140 AA.  
XX  
AC ABU01020;  
XX  
DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX  
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KW gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
FN WO200277021-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-IB002163.  
XX  
PR 27-MAR-2001; 2001GB-00007658.  
XX  
XX (CHIR-) CHIRON SPA.  
XX (GENO-) INST GENOMIC RES.  
XX  
PI Maignani V, Tettelin H, Fraser C;  
XX  
XX WPI; 2003-040579/03.  
DR N-PSDB; ABX06302.  
XX  
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
XX Claim 1; SEQ ID NO 1180; 56pp; English.  
XX  
XX The invention relates to a protein comprising or having at least 50%  
XX identity to any of the 2469 amino acid sequences, identified in the  
XX specification (available on a computer readable format), or its fragment,  
XX expressed from 2469 of 2489 identified DNA coding regions from the  
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
XX AB556454. Also included are an antibody which binds one of the proteins,  
XX treating a patient by administering the protein, DNA or antibody (in a  
XX composition), a kit comprising first and second primers, which are the  
XX nucleic acid cited above or fragments between nucleotides 8-100 of a  
XX sequence not defined in the specification, for amplifying a target

CC sequence contained within a Streptococcus nucleic acid sequence, where  
CC the first primer is substantially complementary to the target sequence  
CC and the second primer is substantially complementary to the complement of  
CC the target sequence, and where the parts of the primers having  
CC substantial complementarity define the termini of the target sequence to  
CC be amplified, assay comprising contacting a test compound with the  
CC protein, and determining whether the test compound binds to the protein  
CC and a Streptococcus pneumoniae bacterium, where one or more genes  
CC encoding the proteins has been rendered inactive. The proteins, nucleic  
CC acid molecules, antibody and compositions are useful as medicaments for  
CC treating or preventing a disease or infection due to Streptococcus  
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
CC media or ear infection. They are also useful in developing vaccines,  
CC diagnostics and antibiotics. The methods are useful for identifying  
CC immunodominant proteins. The present sequence is one of the 2469 proteins  
CC expressed by the identified coding regions from the genomic sequence.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX

SQ Sequence 2140 AA;

Query Match 100.0%; Score 848; DB 6; Length 2140;

Best Local Similarity 100.0%; Pred. No. 1.2e-73;

Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60

Db 1943 TTVKEFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2002

QY 61 WEISGFEKKDAGYVNLNLSKDTFKIPVKPIEKKKEENKPTFDVSKKKNPQVNHSQLN 120

Db 2003 WEISGFEKKDAGYVNLNLSKDTFKIPVKPIEKKKEENKPTFDVSKKKNPQVNHSQLN 2062

QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 164

Db 2063 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2106

RESULT 4

ABU45746

ID ABU45746 standard; protein; 2140 AA.

XX AC ABU45746;

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #31273.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Streptococcus pneumoniae.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio J, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA49616.

XX

PT

PT

PT

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Query Match 100.0%; Score 848; DB 6; Length 2140;

Best Local Similarity 100.0%; Pred. No. 1.2e-73;

Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60

Db 1943 TTVKEFILNKDTGEVSELKPHRVTTVTTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2002

QY 61 WEISGFEKKDAGYVNLNLSKDTFKIPVKPIEKKKEENKPTFDVSKKKNPQVNHSQLN 120

Db 2003 WEISGFEKKDAGYVNLNLSKDTFKIPVKPIEKKKEENKPTFDVSKKKNPQVNHSQLN 2062

QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 164

Db 2063 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2106

RESULT 5

ADM92113

ID ADM92113 standard; protein; 2140 AA.

XX AC ADM92113;

XX 03-JUN-2004 (first entry)

XX S pneumoniae antigenic protein sequence SeqID310.

XX antibacterial; gene therapy; Streptococcus pneumoniae infection;

XX antigenic.

XX Streptococcus pneumoniae.

XX

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 73670; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway of identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 2140 AA;

PN WO2004020609-A2.  
 XX 11-MAR-2004.  
 XX PF 02-SEP-2003; 2003WO-US027401.  
 XX PR 30-AUG-2002; 2002US-0407082P.  
 XX PA (TUFT ) UNIV TUFTS.  
 XX PI Camilli A, Hava DL;  
 XX WPI; 2004-239189/22.  
 XX DR N-PSDB; ADM91876.  
 XX PT New Streptococcus pneumoniae nucleic acid molecules, useful for  
 PT diagnosing, treating and preventing active infections of Streptococcus  
 PT pneumoniae.  
 XX PS Claim 27; SEQ ID NO 310; 123pp; English.  
 XX CC This invention relates to novel isolated Streptococcus pneumoniae nucleic  
 CC acid molecules and the antigenic polypeptides encoded by them. The  
 CC invention may be useful for the production of compounds with an  
 CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
 CC compositions and methods disclosed are useful for treating Streptococcus  
 CC pneumoniae infection. The present sequence is that of an S pneumoniae  
 CC protein of the invention.  
 XX SQ Sequence 2140 AA;

Query Match 100.0%; Score 848; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-73;  
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60  
 DB 1943 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 2002

QY 61 WEISGPEGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSKKKONPQVNHSQLN 120  
 DB 2003 WEISGPEGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSKKKONPQVNHSQLN 2062

QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 164  
 DB 2063 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106

RESULT 6  
 ADP50099  
 ID ADT50099 standard; protein; 2140 AA.  
 XX AC ADT50099;  
 XX DT 13-JAN-2005 (first entry)  
 XX DE S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
 XX KW hyperimmune serum reactive antigen; antibacterial; vaccine;  
 KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
 KW sepsis; meningitis.  
 XX OS Streptococcus pneumoniae TIGR4.  
 XX WO2004092209-A2.  
 XX PD 28-OCT-2004.  
 XX PF 15-APR-2004; 2004WO-BP003984.  
 XX PR 15-APR-2003; 2003BP-00450087.  
 XX PA (INTE-) INTERCELL AG.

XX Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;  
 XX WPI; 2004-758335/74.  
 XX DR N-PSDB; ADT49955.  
 XX PT New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
 PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
 PT treating S. pneumoniae infections.  
 XX PS Disclosure; SEQ ID NO 177; 191pp; English.  
 XX CC This invention relates to novel nucleic acids encoding hyperimmune serum  
 CC reactive antigens, or fragments derived thereof. Specifically, it refers  
 CC to antigens selected from peptides and serum reactive epitopes that can  
 CC be used in pharmaceutical compositions that exhibit antibacterial  
 CC activity. The present invention describes a composition (including the  
 CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
 CC that is useful for manufacturing a medicament such as a vaccine, which  
 CC can be used to treat or prevent bacterial infections, particularly S.  
 CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
 CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
 CC be used for isolating, purifying and/ or identifying an interaction  
 CC partner of the hyperimmune serum reactive antigen, as well as for  
 CC manufacturing a functional nucleic acid selected from aptamers and  
 CC spiegelmers or for manufacturing a functional ribonucleic acid selected  
 CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
 CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
 CC of the invention.  
 XX SQ Sequence 2140 AA;

Query Match 100.0%; Score 848; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-73;  
 Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60  
 DB 1943 TTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDG 2002

QY 61 WEISGPEGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSKKKONPQVNHSQLN 120  
 DB 2003 WEISGPEGKDGAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSKKKONPQVNHSQLN 2062

QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 164  
 DB 2063 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2106

RESULT 7  
 ADR94534  
 ID ADR94534 standard; protein; 637 AA.  
 XX AC ADR94534;  
 XX DT 16-DEC-2004 (first entry)  
 XX DE Novel S. pneumoniae protein sequence, SEQ ID 3169.  
 XX KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
 KW bacterial infection.  
 XX OS Streptococcus pneumoniae.  
 XX US6800744-B1.  
 XX PD 05-OCT-2004.  
 XX PF 30-JUN-1998; 98US-00107433.  
 XX PR 02-JUL-1997; 97US-0051553P.  
 XX PR 12-MAY-1998; 98US-0085131P.  
 XX PA



PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2004-697205/68.  
DR N-PSDB; ADR91931.  
XX  
PT New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide, useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 3169; 151pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366 polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridizable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed S.  
CC pneumoniae protein sequences. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.  
XX  
SQ Sequence 637 AA;  
Query Match 99.6%; Score 845; DB 8; Length 637;  
Best Local Similarity 99.4%; Pred. No. 4.3e-74;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTVKEFILNKDTGEVSELKPHRVTTVTTQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 60  
Db |||||  
QY 61 WEISGFEKGDAGYVNLKDTFIKPVFKIEEKKKEENKPTFDVSKKONPQVNHSQLN 120  
Db |||||  
QY 500 WEISGFEKGDAGYVNLKDTFIKPVFKIEEKKKEENKPTFDVSKKONPQVNHSQLN 559  
QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164  
Db |||||  
QY 560 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 603  
Db |||||  
RESULT 8  
ID AEA58404 standard; protein; 637 AA.  
AC AEA58404;  
XX  
XX 25-AUG-2005 (first entry)  
DT  
DE Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:3169.  
XX  
XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;  
KW vaccine.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US2005136404-A1.  
XX

PD 23-JUN-2005.  
XX  
PF 10-JUL-2003; 2003US-00617320.  
XX  
PR 02-JUL-1997; 97US-0051553P.  
PR 12-MAY-1998; 98US-0085131P.  
PR 30-JUN-1998; 98US-00107433.  
XX  
PA (DOUC/) DOUCETTE-STAMM L A.  
PA (BUSH/) BUSH D.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2005-477576/48.  
DR N-PSDB; AEA55801.  
XX  
PT New isolated nucleic acid molecules and encoded polypeptides useful for  
PT diagnosing, preventing or treating bacterial infections, particularly  
PT Streptococcus pneumoniae infection.  
XX  
PS Claim 5; SEQ ID NO 3169; 144pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule for detecting,  
CC preventing or treating pathological conditions resulting from bacterial  
CC infection. The isolated nucleic acid comprises: (a) any of the 2603  
CC nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence  
CC encoding a Streptococcus pneumoniae polypeptide comprising any of the  
CC 2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide  
CC sequence of at least 8 nucleotides in length, where the sequence is  
CC hybridizable to a nucleic acid having any of the nucleotide sequences in  
CC (a). Also described: (1) a recombinant expression vector comprising the  
CC above nucleic acid operably linked to a transcription regulatory element;  
CC (2) a cell comprising the recombinant expression vector; (3) producing an  
CC S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence  
CC consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)  
CC treating a subject for S. pneumoniae infection; (6) a recombinant or  
CC substantially pure preparation of an S. pneumoniae polypeptide or its  
CC fragment, where the polypeptide is selected from AEA57839 to AEA60441;  
CC (7) a vaccine composition for preventing or treating an S. pneumoniae  
CC infection, comprising an amount of the above nucleic acid or polypeptide;  
CC (8) detecting the presence of a Streptococcus nucleic acid in a sample;  
CC (9) a computer readable medium having recorded the nucleotide sequences  
CC of AEA55236 to AEA57838; (10) a computer based system for identifying  
CC fragments of the Streptococcus genome of commercial importance. The  
CC composition and methods are useful for diagnosing, preventing or treating  
CC bacterial infections, particularly S. pneumoniae infection. The present  
CC sequence represents a S. pneumoniae ORF amino acid sequence from the  
CC present invention. Note - The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from the USPTO web site.  
XX  
SQ Sequence 637 AA;  
Query Match 99.6%; Score 845; DB 9; Length 637;  
Best Local Similarity 99.4%; Pred. No. 4.3e-74;  
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTVKEFILNKDTGEVSELKPHRVTTVTTQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 60  
Db |||||  
QY 440 TTVKEFILNKDTGEVSELKPHRVTTVTTQNGKMSSTIVSEEDFILPVYKGELEKGYQPDG 499  
QY 61 WEISGFEKGDAGYVNLKDTFIKPVFKIEEKKKEENKPTFDVSKKONPQVNHSQLN 120  
Db |||||  
QY 500 WEISGFEKGDAGYVNLKDTFIKPVFKIEEKKKEENKPTFDVSKKONPQVNHSQLN 559  
QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164  
Db |||||  
QY 560 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 603  
Db |||||  
RESULT 9  
ID ADK48759 standard; protein; 2138 AA.  
XX

XX ADK48759;  
 AC 20-MAY-2004 (first entry)  
 DT Streptococcus pneumoniae protein, Seq ID No 5274.  
 DE Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.  
 XX Streptococcus pneumoniae.  
 OS US6699703-B1.  
 PN 02-MAR-2004.  
 PD 26-MAY-2000; 2000US-00593110.  
 PF 02-JUL-1997; 97US-0051553P.  
 PR 12-MAY-1998; 98US-0085131P.  
 PR 30-JUN-1998; 98US-00107433.  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;  
 XX WPI; 2004-212399/20.  
 XX DR N-PSDB; ADK46098.  
 XX New nucleic acid molecules and polypeptides useful for diagnosing,  
 PT preventing and treating pathological conditions resulting from bacterial  
 PT infection, e.g. Streptococcus pneumoniae infection, and in drug  
 PT screening.  
 XX Disclosure; SEQ ID NO 5274; 301pp; English.  
 XX The invention relates to isolated Streptococcus pneumoniae nucleic acids  
 CC and polypeptides. The nucleic acids and proteins are useful for  
 CC diagnosing, preventing and treating pathological conditions resulting  
 CC from bacterial infection, such as S. pneumoniae infection. These may also  
 CC be used for drug screening procedures. The present sequence represents a  
 CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence  
 CC data for this patent did not appear in the printed specification but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html.  
 XX seqdata.uspto.gov/sequence.html.  
 XX Sequence 2138 AA;  
 SQ  
 Query Match 99.6%; Score 845; DB 8; Length 2138;  
 Best Local Similarity 99.4%; Pred. No. 2.3e-73;  
 Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TTVKEPILKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 60  
 DB 1941 TTVKEPILKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDG 2000  
 QY 61 WEISGPEGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKDNQVNHSQLN 120  
 DB 2001 WEISGPEGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSKKDNQVNHSQLN 2060  
 QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164  
 DB 2061 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2104  
 RESULT 10  
 AAW55096  
 ID AAW55096 standard; protein; 117 AA.  
 XX AAW55096;  
 AC AAW55096;  
 XX 02-OCT-1998 (first entry)  
 DT Streptococcus pneumoniae SP0043 protein.  
 XX Streptococcus pneumoniae.

XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
 KW detection; pneumonia; otitis media; meningitis.  
 XX Streptococcus pneumoniae.  
 OS WO9818930-A2.  
 PN 07-MAY-1998.  
 PD 30-OCT-1997; 97WO-US019422.  
 PF 31-OCT-1996; 96US-0029960P.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 PA Kunsch CA, Choi GH, Johnson LS, Hromockyj A;  
 XX WPI; 1998-272224/24.  
 XX DR N-PSDB; AAV27357.  
 XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae  
 PT - or their epitope-containing fragments, useful in protective or  
 PT therapeutic vaccines, and for diagnosis.  
 XX Claim 11; Page 62; 118pp; English.  
 PS The present sequence represents a protein from Streptococcus pneumoniae.  
 XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
 CC can be useful in vaccines for inducing protective antibodies against  
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
 CC are used to detect Streptococcus infection (by usual hybridisation or  
 CC amplification methods), also for isolating Streptococcus genes or their  
 CC allelic variants. The protein can be used similarly to detect specific  
 CC antibodies in standard immunoassays, especially for diagnosing or  
 CC monitoring infections. Antibodies which bind the protein are used to  
 CC detect corresponding antigens, to purify the protein and for passive  
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
 CC (especially 10-300) mu g/ml per dose  
 XX Sequence 117 AA;  
 SQ  
 Query Match 72.5%; Score 615; DB 2; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-52;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 48 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSK 107  
 DB 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVNLKDTFIKPVFKKIEKKKEENKPTFDVSK 60  
 QY 108 KDNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164  
 DB 61 KDNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117  
 RESULT 11  
 ABP54590  
 ID ABP54590 standard; protein; 117 AA.  
 XX ABP54590;  
 AC ABP54590;  
 XX 04-SEP-2002 (first entry)  
 DT S. pneumoniae SP043 protein sequence SEQ ID NO:68.  
 DE Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
 KW antibacterial; Streptococcal infection; detection.  
 XX Streptococcus pneumoniae.  
 OS US2002061545-A1.  
 PN

```

XX PD 23-MAY-2002.
XX PF 22-JAN-2001; 2001US-00765272.
XX XX 30-OCT-1997; 97US-00961083.
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2002-479261/51.
XX DR N-PSDB; ABQ84825.
XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX PT and for preventing or attenuating disease caused by Streptococcus
XX PT infection.
XX PS Claim 11; Page 29; 70pp; English.
XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX CC pneumoniae antigens have antibacterial activity and can be used in
XX CC vaccines. The S. pneumoniae antigens can also be used to prevent or
XX CC attenuate a Streptococcal infection in an animal. The polynucleotides
XX CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX CC of S. pneumoniae ORFs (open reading frames) which are used in an example
XX CC from the present invention
XX SQ Sequence 117 AA;
XX Query Match 72.5%; Score 615; DB 5; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-52;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 48 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 107
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 60
QY 108 KDNFQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 164
Db 61 KDNFQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 117
XX RESULT 12
XX ADC45149
XX ID ADC45149 standard; protein; 117 AA.
XX AC ADC45149;
XX DT 18-DEC-2003 (first entry)
XX DE S. pneumoniae antigenic protein SP043.
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX OS Streptococcus pneumoniae.
XX XX US6573082-B1.
XX PN 03-JUN-2003.
XX PD 28-MAR-2000; 2000US-00536784.
XX PF 31-OCT-1996; 96US-0029960P.
XX PR

```

```

PR 30-OCT-1997; 97US-00961083.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2003-764574/72.
XX DR N-PSDB; ADC45148.
XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX PT useful for producing vaccines for prevention or attenuation of infection
XX PT by Streptococcus pneumoniae.
XX PS Example 1; SEQ ID NO 68; 56pp; English.
XX CC The invention relates to an isolated polynucleotide consisting of a
XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX CC antigens. Also included are making a recombinant vector by inserting the
XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at
XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae
XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX CC antigen nucleic acids are useful as probes for use in diagnostic methods
XX CC for detecting S. pneumoniae gene expression. The present sequence
XX CC represents an S. pneumoniae antigenic protein.
XX SQ Sequence 117 AA;
XX Query Match 72.5%; Score 615; DB 7; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-52;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 48 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 107
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 60
QY 108 KDNFQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 164
Db 61 KDNFQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 117
XX RESULT 13
XX AAG81779
XX ID AAG81779 standard; protein; 746 AA.
XX AC AAG81779;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.
XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
XX KW endocarditis.
XX OS Staphylococcus epidermidis.
XX XX WO200134809-A2.
XX PN 17-MAY-2001.
XX PD 09-NOV-2000; 2000WO-US030782.
XX PF 09-NOV-1999; 99US-0164258P.
XX PR (GLAX ) GLAXO GROUP LTD.
XX PA Kimmerly WJ;
XX PI WPI; 2001-316495/33.
XX DR N-PSDB; AAH52629.

```









Query Match 13.0%; Score 110.5; DB 8; Length 651;  
Best Local Similarity 22.8%; Pred. No. 0.1;  
Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;  
QY 4 KEFLNKDTGEVSELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 53  
DB 422 EBAVSEKNAHETVE---HEETVQSNEPKADNDGNVSNNSNELNEFEV-----ESE 472  
QY 54 K-----GYQPD-GWEISGF--EGKDGAG-----YVINLSKDTFIKPVFKKIE 92  
DB 473 KSEHARSKAKBASSYDYLWGFEFGGVPEHKKEENMLSHLYVSKDKENISKENDVDLD 532  
QY 93 EKKEENKPTFDVSKKDNQPNVHSQLN-----ESHKEDLQREHSQKSDS 139  
DB 533 E-KEEABETEELKEEKEETESEISEDEEEEEKEEENEKKEQKEQSQNNNDQ 591  
QY 140 TKDVTATVLDKNNISSKSTNN 161  
DB 592 KKDMA-----QNLISKQNNN 608

RESULT 21  
ADO19010  
ID ADO19010 standard; protein; 651 AA.  
XX  
AC ADO19010;  
XX  
DT 12-AUG-2004 (first entry)  
XX  
DE P. falciparum GLURP-MSP3 fusion protein.  
XX  
KW Glutamate-rich protein; GLURP-MSP3 fusion protein;  
KW merozoite surface protein 3; malarial vaccine; malaria; immune response;  
KW antimalarial; immunostimulant.  
XX  
OS Plasmodium falciparum.  
OS Synthetic.  
XX  
PN WO2004043488-A1.  
XX  
PD 27-MAY-2004.  
XX  
PP 06-NOV-2003; 2003WO-DK000759.  
XX  
PR 12-NOV-2002; 2002DK-00001741.  
PR 11-SEP-2003; 2003DK-00001307.  
XX  
PA (STAT-) STATENS SERUM INST.  
XX  
PI Theisen M, Jepsen S;  
XX  
DR WPI; 2004-411650/38.  
DR N-PSDB; ADO19011.  
XX  
PT New antigen based vaccine comprising a fusion protein derived from  
PT Plasmodium falciparum Glutamate-rich protein, useful in treating or  
PT preventing malaria.  
XX  
PS Claim 5; SEQ ID NO 1; 52pp; English.  
XX  
CC The present invention relates to a fusion protein comprising Plasmodium  
CC falciparum glutamate-rich protein (GLURP) coupled to P. falciparum  
CC merozoite surface protein 3 (MSP3). The GLURP-MSP3 fusion protein is  
CC useful as an antigen based vaccine against malaria. Also disclosed is the  
CC polynucleotide sequence encoding the GLURP-MSP3 fusion protein. The  
CC polynucleotide sequence is also useful in preparing a vaccine. The  
CC vaccine is useful in treating and preventing malaria and for inducing an  
CC immune response against malaria. The present sequence represents P.  
CC falciparum GLURP-MSP3 fusion protein.  
XX  
SQ Sequence 651 AA;  
Query Match 13.0%; Score 110.5; DB 8; Length 651;

Best Local Similarity 22.8%; Pred. No. 0.1;  
Matches 46; Conservative 38; Mismatches 59; Indels 59; Gaps 10;  
QY 4 KEFLNKDTGEVSELKPHRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELE 53  
DB 422 EBAVSEKNAHETVE---HEETVQSNEPKADNDGNVSNNSNELNEFEV-----ESE 472  
QY 54 K-----GYQPD-GWEISGF--EGKDGAG-----YVINLSKDTFIKPVFKKIE 92  
DB 473 KSEHARSKAKBASSYDYLWGFEFGGVPEHKKEENMLSHLYVSKDKENISKENDVDLD 532  
QY 93 EKKEENKPTFDVSKKDNQPNVHSQLN-----ESHKEDLQREHSQKSDS 139  
DB 533 E-KEEABETEELKEEKEETESEISEDEEEEEKEEENEKKEQKEQSQNNNDQ 591  
QY 140 TKDVTATVLDKNNISSKSTNN 161  
DB 592 KKDMA-----QNLISKQNNN 608  
RESULT 22  
ADZ79639  
ID ADZ79639 standard; protein; 188 AA.  
XX  
AC ADZ79639;  
XX  
DT 14-JUL-2005 (first entry)  
XX  
DE P. falciparum merozoite surface protein 3, MSP3a to MSP3f fragment.  
XX  
KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
KW immunotherapy; malaria; antimalarial; vaccine.  
XX  
OS Plasmodium falciparum.  
XX  
PN WO2005040206-A1.  
XX  
PD 06-MAY-2005.  
XX  
PF 22-OCT-2004; 2004WO-EP012910.  
XX  
PR 24-OCT-2003; 2003US-00691672.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Druilhe P;  
XX  
DR WPI; 2005-355821/36.  
XX  
CC Chimeric molecule useful for preparing vaccine composition against  
CC malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
CC protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
CC immunized with molecule.  
XX  
PS Disclosure; SEQ ID NO 7; 79pp; English.  
XX  
CC The invention relates to a chimeric molecule that comprises a glutamate-  
CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
CC acid residues 25-514) of GLURP (given as SEQ ID NO:1) and a Merozoite  
CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
CC of MSP3 (given as SEQ ID NO:2), wherein the chimeric molecule raises  
CC antibodies against both polypeptides in mice immunized with it. Also  
CC described are: (i) a conjugate comprising the chimeric molecule of the  
CC invention bound to a solid support, (ii) an immunogenic composition  
CC comprising the chimeric molecule, the conjugate described above, or a  
CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
CC against malaria comprising the chimeric molecule, the conjugate described  
CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
CC association with a suitable vehicle, (iv) use of purified and/or  
CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
CC medicament against malaria, and (v) a medicament for passive  
CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
CC antibodies. The chimeric molecule of the invention or a mixture of GLURP



CC and MSP3 antigens are useful for the preparation of a vaccine composition  
CC against malaria. This sequence represents Plasmodium falciparum MSP3a to  
CC MSP3f fragment. Note: The present sequence given as SEQ ID NO:7 in the  
CC Sequence Listing is not mentioned elsewhere in the specification.  
XX  
SQ Sequence 188 AA;

Query Match 12.7%; Score 108; DB 9; Length 188;  
Best Local Similarity 23.2%; Pred. NO. 0.032;  
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
QY 27 IQNGKEMSTIVSEEDFILPVYKGELEKGYQDGEWISGF--EGKKDAG-----YVINLS 79  
DB 15 VLKAKEASS-----YDYL-----GWFGGVPEHKKERNMLSHLVSKD 55  
QY 80 KDTFKVPFKIEEKKE-----ENKPTFVSKKKDPQVNHSLNHRKE 126  
DB 56 KENISKENDVDLDEKEEAETEESELEKNEETSEISEDEEESEEEKEEENDKKK 115  
QY 127 DLQREHSQKSDTKDVTATVLDKNNISKSTNN 161  
DB 116 EQEKEQSNENNQKKDMEA-----QNLISKQNNN 145

RESULT 23  
ADT56185  
ID ADT56185 standard; protein; 470 AA.  
XX  
AC ADT56185;

DT 13-JAN-2005 (first entry)  
XX  
DE Plant polypeptide, SEQ ID 6262.  
XX  
KW Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX  
OS Viridiplantae.

XX  
PN US2004216190-A1.  
XX  
PD 28-OCT-2004.

PF 18-DEC-2003; 2003US-00739930.  
XX  
PR 28-APR-2003; 2003US-00424599.  
PR 28-APR-2003; 2003US-00425115.

XX  
PA (KOVA/) KOVALIC D K.  
XX  
PI Kovalic DK;

XX  
DR WPI; 2004-757369/74.

PT New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.

XX  
PS Claim 2; SEQ ID NO 6262; 14pp; English.

XX  
CC The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed

CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant protein  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.

XX  
SQ Sequence 470 AA;

Query Match 12.7%; Score 107.5; DB 8; Length 470;  
Best Local Similarity 20.1%; Pred. NO. 0.13;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 19 KPHVTVTIOGKEMSTIVSEEDFILPVYKGELEKGYQDGEWISGF-----GKK 70  
DB 82 RENRVTDFVQNNNGESK-----YVODLARRIRYDE-RATGSSQQRIDHPNOK 129

QY 71 DAGVIVNLSKDTFKVPFKIEEKKEENKPTFVSKKKDN----- 111  
DB 130 NVGITERAFENSPIEETSHRVDDNKRINNQNFTAAKSENNAVSRVSGADHKRAEVMGK 189

QY 112 PQVNHSQLNE-----SHRKEDLQREHSQKSDTKDVTATVLDKNNISKSTNNPNK 164  
DB 190 PMENRDQVRQTESAKSHRKENVTKSEKPRDQGVKTEAKDKRNKKEKSEKTESINK 248

RESULT 24  
AAG47777  
ID AAG47777 standard; protein; 484 AA.  
XX  
AC AAG47777;

XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.

XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX  
OS Arabidopsis thaliana.

XX  
PN EP1033405-A2.

XX  
PD 06-SEP-2000.

XX  
PF 25-FEB-2000; 2000EP-00301439.

XX  
PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0126264P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144844P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145152P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 10-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154019P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.



```
XX 30-MAR-2001; 2001US-00820843.
XX PF
XX PR
XX PA
XX (BRAH/) BRAHMACHARI S K.
XX PA (RAMA/) RAMACHANDRAN S.
XX PA (NAND/) NANDI T.
XX PA (BHIM/) BHIMARAO C.
XX
XX Brahmachari SK, Ramachandran S, Nandi T, Bhimarao C;
XX WPI; 2003-492159/46.
XX
XX Identifying candidate proteins useful as anti-infectives involves
XX matching outlier protein sequences with protein sequences in databases.
XX
XX Example 7; Page 91-93; 117pp; English.
XX
XX The present invention relates to a method for identifying candidate
XX proteins in pathogens useful as anti-infectives. The invention discloses
XX a computational method which involves the calculation of several sequence
XX attributes and their subsequent analysis results in the identification
XX of outlier proteins in different pathogens. The method is useful for the
XX identification of outlier proteins (e.g. virulence proteins, antigens or
XX proteins used as drug targets) in pathogenic organisms. The method of the
XX invention provides reproducible results as it does not depend on the
XX variable biochemical characterisation of proteins. AB023500-AB023617
XX represent outlier proteins identified from different pathogenic organisms
XX
XX Sequence 665 AA;
XX
XX Query Match 12.5%; Score 106; DB 7; Length 665;
XX Best Local Similarity 24.3%; Pred. No. 0.29;
XX Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;
XX
XX QY 10 KDTGVSSELKHURVT-VTQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEI--SGF 66
XX DB 127 EEKNKINKSDLRQHNLNLSQSK-----NQDI-----NKNKGKQ-----DISNSNA 169
XX
XX QY 67 EGKDGAGVINLSKDTFKPVFKIEEKB-----EENKPTFD-----VSKKKDNP 112
XX DB 170 ENKCD-----VKEGVKELEEKKEEKISDDHKVEENKKSDDHKVEENKKSDDH 217
XX
XX QY 113 QYNHQLNESHRKEDLQR-BEHSQKSDSTKYDTATVLDKNNISSKSTTNNPNK 164
XX DB 218 KVEENKSDDKHIEVKYVEEHEDEE-----DKKEKKSENKKNKDNK 261
XX
XX
XX RESULT 27
XX ADZ79634
XX ID ADZ79634 standard; protein; 169 AA.
XX AC
XX AC ADZ79634;
XX
XX 14-JUL-2005 (first entry)
XX
XX DE P. falciparum merozoite surface protein 3, amino acid residues 212-380.
XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;
XX KW immunotherapy; malaria; antimalarial; vaccine.
XX
XX OS Plasmodium falciparum.
XX
XX FH Key Location/Qualifiers
XX FT Region 1..169
XX FT /note= "Amino acid residues 212-380 of MSP3"
XX
XX PN WO2005040206-A1.
XX
XX 06-MAY-2005.
XX
XX 22-OCT-2004; 2004WO-EP012910.
```

```
XX 24-OCT-2003; 2003US-00691672.
XX (INSP ) INST PASTEUR.
XX
XX Druilhe P;
XX
XX WPI; 2005-355821/36.
XX
XX Chimeric molecule useful for preparing vaccine composition against
XX malaria, comprises glutamate-rich protein GLURP and Merozoite surface
XX protein 3 MSP3 moieties, and raises antibodies against moieties in mice
XX immunized with molecule.
XX
XX Claim 2; SEQ ID NO 2; 79pp; English.
XX
XX The invention relates to a chimeric molecule that comprises a glutamate-
XX rich protein (GLURP) moiety consisting of a polypeptide fragment (amino
XX acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite
XX surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380
XX of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises
XX antibodies against both polypeptides in mice immunized with it. Also
XX described are: (i) a conjugate comprising the chimeric molecule of the
XX invention bound to a solid support, (ii) an immunogenic composition
XX comprising the chimeric molecule, the conjugate described above, or a
XX mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine
XX against malaria comprising the chimeric molecule, the conjugate described
XX above, or a mixture of GLURP and MSP3 antigens as an immunogen, in
XX association with a suitable vehicle, (iv) use of purified and/or
XX recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a
XX medicament against malaria, and (v) a medicament for passive
XX immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP
XX antibodies. The chimeric molecule of the invention or a mixture of GLURP
XX and MSP3 antigens are useful for the preparation of a vaccine composition
XX against malaria. This sequence represents Plasmodium falciparum MSP3
XX protein (amino acid residues 212-390).
XX
XX Sequence 169 AA;
XX
XX Query Match 12.2%; Score 103.5; DB 9; Length 169;
XX Best Local Similarity 25.2%; Pred. No. 0.077;
XX Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;
XX
XX QY 31 KEMSGTIVSEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 83
XX DB 1 KEASS-----YDYIL-----GMEFGGVPEHKKEENMLSHLYVSSKDKENI 41
XX
XX QY 84 IKPVFKKIEEKEENKPTFDVSKKKNPQVNHSQLN-----ESHKREDLQR 130
XX DB 42 SKENDDVLDL-KEEAEETESEELEEKNEETETSEISEDEEEEEEEKEEKEEKEQEK 100
XX
XX QY 131 BEHSQKSDSTKYDTATVLDKNNISSKSTTNN 161
XX DB 101 EQSNNENDQKKDEA-----QNLISKNQNNN 126
XX
XX
XX RESULT 28
XX ABU24404
XX ID ABU24404 standard; protein; 903 AA.
XX AC
XX AC ABU24404;
XX
XX 19-JUN-2003 (first entry)
XX
XX DE Protein encoded by Prokaryotic essential gene #9931.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Clostridium botulinum.
XX
XX PN WO200277183-A2.
XX
XX 03-OCT-2002.
```



ADP25441 standard; protein; 1791 AA.

ADP25441; 11.8%; Score 100; DB 8; Length 1791; Best Local Similarity 24.2%; Pred. No. 4.5; Matches 52; Conservative 34; Mismatches 71; Indels 58; Gaps 9

09-SEP-2004 (first entry)

Plasmodium falciparum antigen amino acid sequence SEQ ID NO:18.

Plasmodium falciparum; malaria parasite; antigen; immunogenic; immune response; cytosolic; anti-HIV; virucide; hepatotropic; antibacterial; vaccine; cancer; infectious disease; AIDS; hepatitis; bacterial infection.

Plasmodium falciparum.

W02004053086-A2.

24-JUN-2004.

08-DEC-2003; 2003WO-US038966.

06-DEC-2002; 2002US-0431494P.

(EPIM-) EPIMMUNE INC. (USNA ) US SEC OF NAVY.

Sette A, Doolan DL, Carucci DJ, Sidney J, Southwood S; WPI; 2004-468856/44.

New isolated and/or purified Plasmodium falciparum polynucleotide sequences, useful in inducing an immune response for preventing and/or treating cancer and infectious diseases, such as AIDS, hepatitis, and bacterial infections.

Claim 22; SEQ ID NO 18; 253pp; English.

The present invention describes an isolated and/or purified Plasmodium falciparum (malaria parasite) antigen polynucleotide sequence, encoding an immunogenic peptide. Also described: (1) a primer or detection probe for hybridisation with a target sequence or the amplicon generated from a target sequence comprising a sequence of at least 8-30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 consecutive nucleotides of any of the polynucleotide sequences as described above; (2) a DNA chip comprising any of the polynucleotide sequences described above; (3) a vector comprising a promoter operably linked to any of the nucleic acid sequences described above; (4) a host cell transformed by the vector of (3) or the polynucleotide described above; (5) a composition comprising a carrier and the polynucleotide described above; (6) a method of inducing an immune response in an individual comprising the administration of the composition of (5) to induce an immune response; (7) an isolated polypeptide comprising any of the amino acid sequences as encoded by the polynucleotide described above; (8) a composition comprising a carrier and the polypeptide of (7); (9) a method of detecting P. falciparum in biological samples, comprising contacting a biological sample with the isolated polynucleotide and detecting the hybridisation of the isolated polynucleotides with nucleic acids contained in the sample; (10) a method for eliciting an immune response in an individual, comprising the administration of a composition comprising the polypeptides of (7) to an individual to induce an immune response in the individual; (11) an antibody that specifically binds to the P. falciparum polypeptide of (7); and (12) detecting P. falciparum antigens, comprising contacting a sample from a subject with the polypeptide of (7) and detecting the presence of an antigen-antibody complex or detecting the stimulation of T-cells in the sample. The P. falciparum antigens and immunogenic peptides have cytosolic, anti-HIV, virucide, hepatotropic and antibacterial activities, and can be used in vaccines. The methods and compositions of the present invention are useful for inducing an immune response for the prevention and/or treatment of cancer and infectious diseases, such as AIDS, hepatitis, and bacterial infections. The present sequence represents a P. falciparum antigen amino acid sequence, which is used in the exemplification of the present invention.

[illegible]







CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1184 AA;

Query Match 11.5%; Score 97.5; DB 6; Length 1184;  
Best Local Similarity 26.3%; Pred. No. 4.5;  
Matches 46; Conservative 30; Mismatches 70; Indels 29; Gaps 9;  
QY 14 EVSELKPHRVTTIQ-NGKEMSTIV--SEDPF--ILPVYKGLSKGVQPDGWEISGFEG 68  
Db 240 ELSEVNEHRKVIEKELNEKEQKNVVEKKQEDINKEVEVLQDVIEKSDYIN-SIKGVIS 298  
QY 69 KKDAGVIVNLSKDTF-----IKPVFKIEEKK-----EENKPTFDVSKKKD 110  
Db 299 KKESS--QINLIERIRNFTNEISRKNLEIKDKEKLNENKQYIKLESNK--LSGSEELS 354  
QY 111 NPQVNHSQLNESHKEDLQREHSHSQSDSKDVTATVLD-KNNISSKSTNNPNK 164  
Db 355 TLQENIKVLEGGKDKQKIKLESINNEIELLKESIIDLNNKKQEFNSKLSLTNANK 409

RESULT 37  
ABR64281  
ID ABR64281 standard; protein; 2468 AA.

XX ABR64281;

XX 15-OCT-2003 (first entry)

XX Angiogenesis protein BNO382.

XX Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;  
KW antipsoriatic; antiarteriosclerotic; cardiant; vasotropic; angiogenesis;  
KW gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;  
KW diabetic retinopathy; cardiovascular disease; atherosclerosis;  
KW ischemic limb disease; coronary artery disease.

XX Homo sapiens.

XX WO2003027285-A1.

XX 03-APR-2003.

XX 19-SEP-2002; 2002WO-AU001282.

XX 27-SEP-2001; 2001AU-00007973.

XX 27-SEP-2001; 2001AU-00007974.

XX 11-OCT-2001; 2001AU-00008210.

XX 29-OCT-2001; 2001AU-00008532.

XX 13-NOV-2001; 2001AU-00008838.

XX 28-AUG-2002; 2002AU-00951032.

XX (BION-) BIONOMICS LTD.

XX Gamble JR, Hahn CN, Vadas MA;

XX WPI; 2003-354655/33.

XX N-PSDB; ACF34559.

XX New angiogenic genes and polypeptides, useful for diagnosing, e.g.  
PT prognosticating or treating an angiogenesis-related disorder,  
PT cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or  
XX cardiovascular diseases.

XX Claim 15; SEQ ID NO 216; 90pp; English.

CC The invention relates to the isolation of novel genes (ACF34446-ACF34559)  
CC encoding proteins (ABR64180-ABR64281) involved in the process of  
CC angiogenesis. The nucleic acid molecules are useful in identifying and/or  
CC obtaining full-length human genes involved in an angiogenic process. The  
CC nucleic acid molecule, polypeptides or complexes encoded, cells or  
CC genetically modified non-human animals derived from these are useful for  
CC the screening of candidate pharmaceutical compounds used in treating  
CC angiogenesis-related disorders. They are also useful for diagnosing,  
CC prognosticating or treating an angiogenesis-related disorder, which  
CC involves uncontrolled or enhanced angiogenesis or is a disorder in which  
CC a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,  
CC diabetic retinopathy, psoriasis or cardiovascular diseases such as  
CC atherosclerosis), or involves inappropriately arrested or decreased  
CC angiogenesis or is a disorder in which an expanding vasculature is of  
CC benefit (e.g. ischemic limb disease or coronary artery disease). The  
CC modulator of expression or activity of the polypeptide encoded by the  
CC nucleic acid sequence is useful for manufacturing a medicament for the  
CC treatment of an angiogenesis-related disorder. This sequence corresponds  
CC to one of the novel angiogenic protein  
XX  
SQ Sequence 2468 AA;

Query Match 11.5%; Score 97.5; DB 6; Length 2468;  
Best Local Similarity 24.8%; Pred. No. 12;  
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEFILNKDTGVSSELKPHRVTTTQNGKEMSTIVSEDFILPVYKGELEKGYQFDGWEI 63  
Db 584 EKVMVKDKPKVTETKPSVTEKVPSEEPS-----PV-KAEVA-----EK 623

QY 64 SGFEGKKDAGYVIVNLSKDTFIKVPFKIEEKKEBENKPTPDVSKKONPQVNHSQLNESH 123  
Db 624 QATDVKPKAAKEKTVKKEKVPK-----EDKKEEKKPKKEVAKEDKTPi---KKEKP 675

QY 124 RKEDLQRE-----BHSQKSDSTKDV 143

Db 676 KKEEVKKEVKKEIKKEEKKPKKEV 700

RESULT 38

ADBE62723  
ID ADBE62723 standard; protein; 2468 AA.

XX ADBE62723;

XX 29-JAN-2004 (first entry)

XX Human Protein NP\_005900, SEQ ID NO 8656.

XX Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.

XX Homo sapiens.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

XX 01-NOV-2001; 2001US-0346382P.

XX 26-NOV-2001; 2001US-0333347P.

XX (GEO) GEN HOSPITAL CORP.

XX (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-268312/26.

XX GENBANK; NP\_005900.



DE Human Protein AAA18904, SEQ ID NO 8660.  
XX  
KW Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
XX WPI; 2003-268312/26.  
DR GENBANK; AAA18904.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2468 AA;

Query Match 11.5%; Score 97.5; DB 7; Length 2468;  
Best Local Similarity 24.8%; Pred. No. 12;  
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;  
QY 4 KEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEI 63  
DB 584 EKVMVKDKPKVTEKPSVTEKEVPSKEPS-----PV-KAEVA-----EK 623  
QY 64 SGFEKGKAGYVINLSKDTFIKPVFKIEKKEENKPTFDVSKKKNPQVNHQNLNESH 123  
DB 624 QATDVKPKAAKKTVKETKVKP-----EDKKEEKEKPKKEVAKKEDKTPI----KKEEKP 675

Qy 124 RKEDLORE-----EHSOKSDSTKDV 143  
Db 676 KGEVKKKEVKKEIKKBEKKEPKKEV 700  
Search completed: April 24, 2006, 14:50:28  
Job time : 103.913 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 75.3154 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773

Perfect score: 651

Sequence: 1 EDFILPVYKLEKGYQFDG.....ATVLDKNNISSKSTTNNPK 124

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt.05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	651	100.0	2119	Q9AHT5_STRPN	Q9Aht5 streptococc
2	651	100.0	2140	Q97RY6_STRPN	Q97ry6 streptococc
3	648	99.5	2144	Q9S4M8_STRPN	Q9S4m8 streptococc
4	648	99.5	2144	Q8DQP7_STRPN	Q8dqp7 streptococc
5	125.5	19.3	300	Q4XUI6_PLACH	Q4xui6 plasmodium
6	112	17.2	361	Q9SP15_PLAPA	Q9sp15 plasmodium
7	110	16.9	346	Q9U0G0_PLAPA	Q9u0g0 plasmodium
8	110	16.9	379	Q25705_PLAPA	Q25705 plasmodium
9	109	16.7	3008	Q81436_PLAF7	Q81436 plasmodium
10	108	16.6	384	Q50VJ0_ENTHI	Q50vj0 entamoeba h
11	106.5	16.4	600	Q77355_PLAF7	Q77355 plasmodium
12	106	16.3	354	Q25995_PLAPA	Q25995 plasmodium
13	106	16.3	354	Q81J55_PLAF7	Q81j55 plasmodium
14	106	16.3	616	Q6BRW2_DBBHA	Q6brw2 debaromyce
15	103	15.8	379	Q9U6C4_PLAPA	Q9u6c4 plasmodium
16	102.5	15.7	380	Q26019_PLAPA	Q26019 plasmodium
17	102	15.7	379	Q25706_PLAPA	Q25706 plasmodium
18	102	15.7	540	Q54MT2_DICDI	Q54mt2 dictyosteli
19	101.5	15.6	382	Q9V7J0_DROME	Q9v7j0 drosophila
20	101.5	15.6	556	Q9V7I9_DROME	Q9v7i9 drosophila
21	101.5	15.6	785	Q9GQ82_DROME	Q9gq82 drosophila
22	101.5	15.6	954	Q6HNR0_BACHK	Q6hnr0 bacillus th
23	101	15.5	296	Q50LX8_ENTHI	Q50lx8 entamoeba h
24	101	15.5	662	Q4YMU4_PLAPA	Q4ymu4 plasmodium
25	100	15.4	1011	Q4Y2I3_PLACH	Q4y2i3 plasmodium
26	100	15.4	1130	Q81JZ4_PLAF7	Q81jz4 plasmodium
27	99.5	15.3	1038	Q90784_CHICK	Q90784 gallus gall
28	99	15.2	211	P91488_CAREL	P91488 caenorhabdi
29	99	15.2	329	Q9NFV9_PLAPA	Q9nfv9 plasmodium
30	99	15.2	437	Q54K26_DICDI	Q54k26 dictyosteli
31	99	15.2	2563	Q813A0_PLAF7	Q813a0 plasmodium

32	98.5	15.1	393	2	Q7RKU2_PLAYO	Q7rku2 plasmodium
33	98.5	15.1	674	2	Q7RLI7_PLAYO	Q7rli7 plasmodium
34	98.5	15.1	827	2	Q55G46_DICDI	Q55g46 dictyosteli
35	98	15.1	951	2	O96229_PLAF7	O96229 plasmodium
36	97.5	15.0	556	2	Q9S893_DROME	Q9s893 drosophila
37	97.5	15.0	1345	1	YH00_YEAST	P38800 saccharomyc
38	97	14.9	895	2	Q4Z7T5_PLABE	Q4z7t5 plasmodium
39	97	14.9	1028	2	Q4Z4Q1_PLABE	Q4z4q1 plasmodium
40	96.5	14.8	905	2	Q6BXE1_DBBHA	Q6bxel debaromyce
41	96	14.7	736	2	Q4YVY2_PLABE	Q4yvy2 plasmodium
42	96	14.7	899	2	Q86J06_DICDI	Q86jq6 dictyosteli
43	96	14.7	1042	2	Q7R8Q7_PLAYO	Q7req7 plasmodium
44	96	14.7	1120	2	Q557C8_DICDI	Q557c8 dictyosteli
45	95.5	14.7	500	2	Q6BGL7_PARTE	Q6bgl7 paramecium

#### ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN  
ID Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
DT 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
DOI=10.1128/IAI.69.3.1593-1598.2001;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
Langermann S., Johnson S., Koenig S;  
RA "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598 (2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 28BT.  
DR MEROPS; S08\_064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Protease.

```

FT NON TER 1 1
SQ SEQUENCE 2119 AA; 238227 MW; 517F9B7F6B960A6A CRC64;

Query Match
Best Local Similarity 100.0%; Score 651; DB 2; Length 2119;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVNLNLSKDTFKPVPFKIEEKEEENK 60
|||||
Db 1962 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVNLNLSKDTFKPVPFKIEEKEEENK 2021
|||||

QY 61 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 120
|||||
Db 2022 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 2081
|||||

QY 121 NPNK 124
|||||
Db 2082 NPNK 2085

RESULT 2
Q97RY6_STRPN
ID Q97RY6_STRPN PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Serine protease, subtilase family.
GN OrderedLocName=SP0641;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.P., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollinghead S.K., Frazer C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007373; AAK74791.1; -; Genomic_DNA.
DR PIR; F95074; F95074.
DR HSSP; P00782; 2S8T.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0004308; F:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR01680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.

```

```

DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240426 MW; FFA4AD8E2938B334 CRC64;

Query Match
Best Local Similarity 100.0%; Score 651; DB 2; Length 2140;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVNLNLSKDTFKPVPFKIEEKEEENK 60
|||||
Db 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGKQDAGYVNLNLSKDTFKPVPFKIEEKEEENK 2042
|||||

QY 61 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 120
|||||
Db 2043 PTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTTN 2102
|||||

QY 121 NPNK 124
|||||
Db 2103 NPNK 2106

RESULT 3
Q984M8_STRPN
ID Q984M8_STRPN PRELIMINARY; PRT; 2144 AA.
AC Q984M8;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase precursor PrtA.
GN Name=prtA;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3.B;
RX MEDLINE=21585565; PubMed=11728722;
RA Bethe G., Nau R., Wellner A., Hakenbeck R., Reinert R.R., Heinz H.P.,
RA Zysk G.;
RT "The cell wall-associated serine protease PrtA: a highly conserved virulence factor of Streptococcus pneumoniae.";
RL FEMS Microbiol. Lett. 205:99-104(2001).
DR EMBL; AF127143; AAD48399.1; -; Genomic_DNA.
DR HSSP; P00782; 2S8T.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR01680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.

```

```

DR PROSITE, PS00138; SUBTILASE_SER; UNKNOWN 1.
KW Cell wall; Signal.
FT SIGNAL 1 19 Potential.
FT CHAIN 20 2144 cell wall-associated serine proteinase
FT PTAA.
SQ SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;

Query Match
Best Local Similarity 99.5%; Score 648; DB 2; Length 2144;
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGLKGYQFDGWEISGFEGKDGAGYVINLSKDTFKVPVKKEEENK 60
DB 1987 EDFILPVYKGLKGYQFDGWEISGFEGKDGAGYVINLSKDTFKVPVKKEEENK 2046

QY 61 PTFDVSKKKDPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSTTN 120
DB 2047 PTFDVSKKKDPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSTTN 2106

QY 121 NPNK 124
DB 2107 NPNK 2110

RESULT 4
Q8DQ7_STRR6
ID Q8DQ7_STRR6 PRELIMINARY; PRT; 2144 AA.
AC Q8DQ7.
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN Name=prtA; OrderedLocustNames=pr0561;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
CC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA DeHoff B.S., Alborn W.B. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Hoskins J., Strem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McEaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE084343; AAK99365.1; -; Genomic_DNA.
DR FIC; A97942; A97942.
DR HSSP; P00782; 2SET.
DR MEROPS; S08.054; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.

```

```

DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN 1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN 1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 1.
KW Cell wall; Complete proteome.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match
Best Local Similarity 99.5%; Score 648; DB 2; Length 2144;
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGLKGYQFDGWEISGFEGKDGAGYVINLSKDTFKVPVKKEEENK 60
DB 1987 EDFILPVYKGLKGYQFDGWEISGFEGKDGAGYVINLSKDTFKVPVKKEEENK 2046

QY 61 PTFDVSKKKDPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSTTN 120
DB 2047 PTFDVSKKKDPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSTTN 2106

QY 121 NPNK 124
DB 2107 NPNK 2110

RESULT 5
Q4XUI6_PLACH
ID Q4XUI6_PLACH PRELIMINARY; PRT; 300 AA.
AC Q4XUI6.
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Hypothetical protein (Fragment).
GN ORFNames=PC000286.03.0;
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolj T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAY1003049; CAH79425.1; -; Genomic_DNA.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00678; WD_REPEATS_2; 4.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
FT NON TER 1
SQ SEQUENCE 300 AA; 34469 MW; 8DE4E512AFB1945 CRC64;

Query Match
Best Local Similarity 19.3%; Score 125.5; DB 2; Length 300;
Matches 33; Conservative 26; Mismatches 31; Indels 25; Gaps 6;

QY 5 LPVYKGLKGYQFDGWEISGFEGKDGAGYVINLSKDTFK-----PVFKKEEENK 56

```

Db 164 LAVYRGHVPAYKI-AWSI-----DNNYIVSCSDSTLKLWLNHLVPLLRKEENAE 215  
Qy 57 ---EENKTFDVSKKKQNPVNHSQLNESHKEDLQREHSQKSDSTKDVATVL 108  
Db 216 QTKDSQK-----NEQKENPQ-NDQPNDEANSEKKKKKNEKNDKTKNKIKTKLL 264

## RESULT 6

Q95P15 PLAP7  
ID Q95P15\_PLAP7 PRELIMINARY; PRT; 361 AA.  
AC Q95P15;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=PVO.  
RX MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;  
RA Hiseida H., Saul A., Reese J.J., Kennedy M.C., Long C.A., Miller L.H.,  
RA Stowers A.W.;  
RT "Merozoite surface protein 3 and protection against malaria in Aotus  
RT nancymai monkeys";  
RL J. Infect. Dis. 185:657-664 (2002).  
DR EMBL; AY044180; AAK94780.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
KW Merozoite.  
FT CHAIN <1> >361 merozoite surface protein 3.  
FT NON\_TER 1  
FT NON\_TER 361  
FT SEQUENCE 361 AA; 41163 MW; 6127A3041587BA7A CRC64;

Query Match 17.2%; Score 112; DB 2; Length 361;  
Best Local Similarity 23.9%; Pred. No. 1.5; Mismatches 48; Indels 42; Gaps 6;  
Matches 37; Conservative 28;  
Qy 8 YKGELEKGYQ-----PD-----GWEISGF--EGKQDAG-----YVI 36  
Db 165 YAGKVKDYERAKNAYQKANQAVLKAEASSYDYLGWFGGVPEHKKEENMLSHLYS 224  
Qy 37 NLSKDTFTKPVFKKIEEKEENKPTFDVSKKQNPVNHSQLNESHKEDLQREHSQK 96  
Db 225 SKKENISKENDVDVLDL-KEERAEETEEELSEKNEETSEISEDEEEEEEKEE 283  
Qy 97 SDSTKDVATVLDKN-----NISKSTTN 121  
Db 284 NDKKKEQEQSNENNNDQKQMEAAQLISKQNNN 318

## RESULT 7

Q9U0G0 PLARE  
ID Q9U0G0\_PLARE PRELIMINARY; PRT; 346 AA.  
AC Q9U0G0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
GN Name=mep3;  
OS Plasmodium reichenowi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5854;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;  
RA Okenu D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
RT Plasmodium reichenowi and Plasmodium falciparum";  
RL Mol. Biochem. Parasitol. 109:185-188 (2000).

DR EMBL; AJ252286; CAB65754.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
KW Merozoite.  
FT NON\_TER 1  
FT NON\_TER 346  
FT SEQUENCE 346 AA; 39127 MW; A804B96BDFAFAP010 CRC64;

Query Match 16.9%; Score 110; DB 2; Length 346;

Best Local Similarity 26.2%; Pred. No. 2; Mismatches 26; Indels 16; Gaps 5;  
Matches 34; Conservative 26;  
Qy 7 YKGELEKGYQFD-GWEISGF--EGKQDAG-----YVINLSKDTFTKPVFKKIEEKEE 58  
Db 184 VLKAKEASSYNYILGWFGGVPEHKKEENMLSHLYSSKDKENISKENDVDVLDL-KEE 242  
Qy 59 NKPTFDVSKKQNPVNHSQLNESHKEDLQREHSQKSDSTKDVATV-----LDKN 111  
Db 243 AEETGEQLEKEEKEETSEINEDEEQEKEEKEEENNDKKQKQAKESQSNQDKEDMEAQ 302  
Qy 112 NISKSTTN 121  
Db 303 NLISKQNNN 312

## RESULT 8

Q25705 PLAP7  
ID Q25705\_PLAP7 PRELIMINARY; PRT; 379 AA.  
AC Q25705;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31 (1997).  
DR EMBL; U08851; AAC47831.1; -; Unassigned\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
DR SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 16.9%; Score 110; DB 2; Length 379;  
Best Local Similarity 23.4%; Pred. No. 2.2; Mismatches 30; Indels 50; Gaps 7;  
Matches 37; Conservative 30;  
Qy 8 YKGELEKGYQ-----PD-----GWEISGF--EGKQDAG-----YVI 36  
Db 185 YAGKVKDYERAKNAYQKANQAVLKAEASSYDYLGWFGGVPEHKKEENMLSHLYS 244  
Qy 37 NLSKDTFTKPVFKKIEEKEENKPTFDVSKKQNPVNHSQLN-----ESH 83  
Db 245 SKKENISKENDVDVLDL-KEERAEETEEELSEKNEETSEISEDEEEEEEKEEENE 303  
Qy 84 RKEDLQREHSQKSDSTKDVATVLDKNISKSTTN 121  
Db 304 KKKQEQEQSNENNNDQKQMEAA-----QLNISKQNNN 336

## RESULT 9

Q81436 PLAP7  
ID Q81436\_PLAP7 PRELIMINARY; PRT; 3008 AA.  
AC Q81436;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein PFE0325w.



```

GN Name=PRE0325w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RP STRAIN=3D7;
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Lark N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.B., Craig A., Newbold C., Barrall B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3D7;
RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrall B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL929351; CAD51431.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 3008 AA; 356025 MW; 60BCBBE15C599B4 CRC64;

Query Match 16.7%; Score 109; DB 2; Length 3008;
Best Local Similarity 32.4%; Pred. No. 24;
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 34 YVINLSK----DTFKPVFKIEEKEEENKPTFDVSKKONPQVNHSHQ---LNESHKRE 86
Db 2310 YDIELSKLEKFGASIGPVFTD-EENKEENKN-EVNKEENKKEENKKEENKVE 2366

QY 87 DLQREH----SQSDSTKVATVLDKNISKK-----STTNPNK 124
Db 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414

RESULT 10
Q50VJ0 ENTHI PRELIMINARY; PRT; 384 AA.
AC Q50VJ0;
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE HMG box protein.
OS ORFNames=188.t00012;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RP STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Lofius B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hart R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoest U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z.,
RA Gullen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,

RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrall B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAF90100585; EAL45607.1; -; Genomic DNA.
SQ SEQUENCE 384 AA; 45464 MW; 207789F65D72B019 CRC64;

Query Match 16.6%; Score 108; DB 2; Length 384;
Best Local Similarity 32.0%; Pred. No. 3.1;
Matches 31; Conservative 18; Mismatches 42; Indels 6; Gaps 2;

QY 22 EISGFEKGKADGYVINLSKDTFKPVFKIEEKEEENKPTFDVSKKONPQVNHSHQ 81
Db 268 EMKKEGKESDKKEDTKDKKKVKSSEKDEIKKEDKKH---EKKEEKEEKKPKPE 323

QY 82 SHRKEDLQREHSHKSDSTKD--VTATVLDKNISKK 116
Db 324 SEKESKKEKHKHKKEDKKKDEEKKKVEDKKSKKQK 360

RESULT 11
O77355 PLAF7 PRELIMINARY; PRT; 600 AA.
AC O77355;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein MAL3P4.20.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RP MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson S., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jassal B., Kyes S., Mclean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,
RA Rutter S., Sulston J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrall B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum.";
RL Nature 400:532-538(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden P., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Lark N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.B., Craig A., Newbold C., Barrall B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
DR EMBL; AL008970; CAAL5610.2; -; Genomic_DNA.
DR PIR; T18467; T18467.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0006397; P:mRNA processing; IEA.

```



```
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicoud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
RA Swennen D., Tekala P., Wesolowski-Louvel M., Weethof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Winkler P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR382136; CAG87226.1; -; Genomic DNA.
DR GO; GO:0016301; F-kinase activity; IEA.
DR InterPro; IPR000749; ATP_gua_Prans.
KW Complete proteome.
SQ SEQUENCE 616 AA; 72143 MW; 884009B2B8B6C3CF CRC64;

Query Match
Best Local Similarity 16.3%; Score 106; DB 2; Length 616;
Matches 43; Conservative 20; Mismatches 49; Indels 28; Gaps 9;

QY 6 PVYKGELEKGYQPDGWEISGFGKKDAGVINLSKDT-FIKPV-----PKKIEKKKEENK 60
DB 87 PVLKERRAKPKITLTDIT-----KDIN-DINFSSDSEHKPIETSKTKKTKTKTK 140

QY 61 PTFDVSKKK-----DNPO--VNHSQINE---SHRKEDLQREHSOK-----SDSTKQVT 104
DB 141 PDLDIGKLERITVADNPDIENHSSEBEIKQRKEKQKQKQKQKQKQKQKQKQKQKQ 200

QY 105 ATVLK-NNISKSTNNPN 123
DB 201 TEQPSPLKNINEKITSNPS 220

RESULT 15
QSU6C4_PLAFA
ID QSU6C4_PLAFA PRELIMINARY; PRT; 379 AA.
AC QSU6C4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Polymorphic antigen.
GN Name=MSP-3;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FC37;
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188190; AAF04099.1; -; Genomic DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;

Query Match
Best Local Similarity 15.8%; Score 103; DB 2; Length 379;
Matches 36; Conservative 30; Mismatches 42; Indels 50; Gaps 7;

QY 8 YKGELEKGYQPDGWEISGFGKKDAGVINLSKDT-FIKPV-----PKKIEKKKEENK 60
DB 185 YAEQVEKDYERAKNAYQKQAVLKAEASSYDYLGWFGGVPPEHKKEENMLSHLYVS 244

QY 37 NLSKDTFTKPKVTKKEEKEENKPTFDVSKKQNPQVNHSLN-----ESH 83
DB 245 SKDKEINIKENDVDLDE-KEEAETETEEEEKKEEKEEKEEKEEKEEKEEKEEKEE 303

QY 84 RKEDLQREHSOKSDSTKQVTATVLKNNISKSTNN 121
DB 304 KKEEKEKQSNENNDDQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 336
```

```
RESULT 16
Q26019_PLAFA
ID Q26019_PLAFA PRELIMINARY; PRT; 380 AA.
AC Q26019;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Polymorphic antigen precursor.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FC37;
RX MEDLINE=95198774; PubMed=7891748; DOI=10.1016/0166-6851(94)00149-9;
RA McColl D.J., Silva A., Foley M., Kun J.P., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.P.;
RT "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites.";
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FC37;
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.P.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L07944; AAC09378.1; -; Genomic DNA.
DR PDB; 1PSM; NMR; @=90-127.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
KW SIGNAL.
FT SIGNAL. 1 25 Potential.
FT CHAIN 26 380 polymorphic antigen.
SQ SEQUENCE 380 AA; 43290 MW; 0386CA1393094C2 CRC64;

Query Match
Best Local Similarity 15.7%; Score 102.5; DB 2; Length 380;
Matches 34; Conservative 28; Mismatches 47; Indels 27; Gaps 6;

QY 7 VYKGELEKGYQPDGWEISGFGKKDAGVINLSKDT-FIKPVFKKIEKKKEE 58
DB 208 VLKAEASSYDYLGWFGGVPPEHKKEENMLSHLYSSKDKENISKENDVDLDE-KEE 266

QY 59 NKPTFDVSKKQNPQVNHSLN-----ESHKEDLQREHSOKSDSTKQVTA 105
DB 267 AEETEEEEKKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE 326

QY 106 TVLDKNNISKSTNN 121
DB 327 -----QNLISKQNNN 337

RESULT 17
Q25706_PLAFA
ID Q25706_PLAFA PRELIMINARY; PRT; 379 AA.
AC Q25706;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Polymorphic antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.P.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08852; AAC47832.1; -; Unassigned DNA.
```

DR InterPro; IPR010784; Merozoite SPAM.  
DR Pfam; PF07133; Merozoite SPAM; 1.  
SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;

Query Match 15.7%; Score 102; DB 2; Length 379;  
Best Local Similarity 22.8%; Pred. No. 8.8;  
Matches 36; Conservative 30; Mismatches 42; Indels 50; Gaps 7;

QY 8 YKGELEKGYQ-----PD---GWEISGF--EGKKDAG-----YVI 36  
Db 185 YAEQVEKDYERAKNAYQKANOAVLKAKBASSYDYLWGVEFGVGVEHKKENMLSHLYVS 244

QY 37 NLSKQTFIKVPFKIEEKEENKPTFDVSKKDNQVNHSQLN-----BSH 83  
Db 245 SKDKENISKENDVDLDE-KEEAETETEEELKEKEETETESISEDEEBEERKEEEND 303

QY 84 RKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 121  
Db 304 KKKEQKEQSNENNQKKDMEA-----QNLISKNNNN 336

RESULT 18  
Q54MT2 DICDI PRELIMINARY; PRT; 540 AA.  
AC Q54MT2;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DE Hypothetical protein.  
GN ORFNames=DDB0186654;  
OS Dictyostellium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.  
OX NCBI\_TaxID=44689;  
RN [1]

RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=AX4;  
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,  
RA Sucgang R., Barriman M., Song J., Olsen R., Szafranski K., Xu Q.,  
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivoiro P.,  
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,  
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,  
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,  
RA Farbrother P., Denaby B., Just E., Morio T., Rost R., Churcher C.,  
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,  
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,  
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,  
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,  
RA Loulsegged H., Mungall K., Oliver K., Price C., Quail M.A.,  
RA Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,  
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,  
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,  
RA Shauly G., Schleicher M., Weinstock G., Rosenthal A., Cox E.C.,  
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzner M., Kay R.R.,  
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;  
RT "The genome of the social amoeba Dictyostellium discoideum";  
RL Nature 0-0(2005).

CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC EMBL; AAFI01000124; EAL64677.1; -; Genomic\_DNA.  
KW Hypothetical protein.

SQ SEQUENCE 540 AA; 60799 MW; 0F48391CB55A35B5 CRC64;

Query Match 15.7%; Score 102; DB 2; Length 540;  
Best Local Similarity 29.4%; Pred. No. 13;  
Matches 25; Conservative 18; Mismatches 38; Indels 4; Gaps 1;

QY 44 IKPVPFKIEEKEENKPTFDVSKKDNQVNHSQLNESH-----RKEDLQREHSQKSDS 99  
Db 435 VRPVPKSSSKRKSEPTTEESKSKKSKSKSDDVEMKEPVPKEEKSKSKKS 494

QY 100 TKDVTATVLDKNNISSKSTNNPNK 124  
RP NUCLEOTIDE SEQUENCE.  
MEDLINE=22426070; PubMed=12537573;

Db 495 SKKEEPVPKEKKSSSKKDKCKK 519  
RESULT 19  
Q9V7J0 DROME PRELIMINARY; PRT; 382 AA.  
AC Q9V7J0; Q9GQ81;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
DE CG8421-PB, isoform B (Aspartyl beta-hydroxylase variant 2).  
GN Name=Asph; ORFNames=CG8421;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]

RP NUCLEOTIDE SEQUENCE.  
MEDLINE=2019606; PubMed=10731132; DOI=10.1126/science.287.5461.2185;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fogle C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston F.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jaitai M., Kalush K., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svizskas R., Testor C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL Science 287:2185-2195(2000).  
RN [2]

RP NUCLEOTIDE SEQUENCE.  
MEDLINE=22426065; PubMed=12537568;

RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svizskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]

RP NUCLEOTIDE SEQUENCE.  
MEDLINE=22426070; PubMed=12537573;

RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnick K.S.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Rettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richner C., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RG Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RG FlyBase;  
RL Submitted (MAR-2005) to the EMBL/GenBank/DBSJ databases.  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20564328; PubMed=10956665; DOI=10.1074/jbc.M006753200;  
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.P.,  
RA Friedman P.A.;  
RT "Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved  
RT isoform of Asph missing the catalytic domain share exons with  
RT junctin.";  
RL J. Biol. Chem. 275:39543-39554(2000).  
DR EMBL; AF003808; AAF58063.2; -; Genomic\_DNA.  
DR EMBL; AF289494; AAG40807.1; -; mRNA.  
DR Ensembl; CG8421; Drosophila melanogaster.  
DR FlyBase; FBgn0034075; Asph.  
DR FlyBase; FBgn0034075; CG8421.  
SQ SEQUENCE 382 AA; 43287 MW; 60B5C03ABEFC6E8B CRC64;

Query Match 15.6%; Score 101.5; DB 2; Length 382;  
Best Local Similarity 24.5%; Pred. No. 9.7;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
QY 1 EDPLPVYKGELEKGVQPDGM-----RLSGPEGKKDAGYVI-----NLSKDTFIK 45  
Db 78 ELDLTPLESRRSK--VFDGWVDEHRDEHDGHDVQPSGEALDDHDEHDDHDEDEE 135  
QY 46 PVFKIEEKKEENKPT-----FVSKKKNPQVNHSLNESHRKEDLPQREHSQKSDS 99  
Db 136 PLTESLELEEESEPTDEPAADVEYEDEDEENNA--GENITAEADAESEEDND 193  
QY 100 TKDVTATVLDKNISKST 118  
Db 194 EGTVEATVEATTEATTEAT 212  
RESULT 20  
Q9V719 DROME PRELIMINARY; PRT; 556 AA.  
AC Q9V719;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG8421-PD, isoform D (CG8421-pe, isoform e).

GN Name=Asph; ORFNames=CG8421;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Brill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieri S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Hernandez J.R., Houck J.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nusskern D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shier B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Turner C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskas R., Tabor P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,  
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celniker S.E.;  
RT "The transposable elements of the Drosophila melanogaster euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnick K.S.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Rettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richner C., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.;  
RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
RT systematic review";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RG Berkeley Drosophila Genome Project;  
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,  
RA Yu C., Rubin G.;  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RG FlyBase;  
RL Submitted (MAR-2005) to the EMBL/GenBank/DBSJ databases.  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20564328; PubMed=10956665; DOI=10.1074/jbc.M006753200;  
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.P.,  
RA Friedman P.A.;  
RT "Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved  
RT isoform of Asph missing the catalytic domain share exons with  
RT junctin.";  
RL J. Biol. Chem. 275:39543-39554(2000).  
DR EMBL; AF003808; AAF58063.2; -; Genomic\_DNA.  
DR EMBL; AF289494; AAG40807.1; -; mRNA.  
DR Ensembl; CG8421; Drosophila melanogaster.  
DR FlyBase; FBgn0034075; Asph.  
DR FlyBase; FBgn0034075; CG8421.  
SQ SEQUENCE 382 AA; 43287 MW; 60B5C03ABEFC6E8B CRC64;



```

RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celnikier S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirekas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289493; RAG40806.1; -; mRNA.
DR EMBL; AE003808; AAM70947.1; -; Genomic DNA.
DR Ensembl; CG8421; Drosophila melanogaster.
DR FlyBase; FBgn0034075; Asph.
DR GO; GO:0030176; C:integral to endoplasmic reticulum membrane; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0004597; F:peptide-aspartate beta-dioxygenase activity; IEA.
DR GO; GO:0018193; P:peptidyl-amino acid modification; IEA.
DR InterPro; IPR007803; Asp Arg Hydrol.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001440; TPR-like helical.
DR InterPro; IPR011990; TPR-like helical.
DR Pfam; PF05118; Asp Arg Hydrol; 1.
DR PROSITE; PS0293; TPR REGION; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD5836F7F1 CRC64;

Query Match 15.6%; Score 101.5; DB 2; Length 785;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

Qy 1 EDFILPVYKGELEKGYQFDGW-----RISGFEKKDAGYVI-----NLKDTTFIK 45
Db 78 EDLDTPLSESFRSK--VFQGVDEHRDHDGHDVQPSGEALDDHDDHDDHDEDEE 135

Qy 46 PVFKKIKKCEENKPT-----PDVSKKNQPNVHSHKEDLQREHSOKSDS 99
Db 136 PLTEELELEHEEEPTEDPADEYEDYEDEENNA--GENITADEEEEDND 193

Qy 100 TKDVTATVLDKNNISKST 118
Db 194 EGTVEATVEATTAT 212

RESULT 22
Q6HNR0 BACHK
ID Q6HNR0 BACHK PRELIMINARY; PRT; 954 AA.
AC Q6HNR0
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Possible internalin protein.
GN OrderedLocustNames=B79727_0463;
OS Bacillus thuringiensis (subsp. konkukian).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group.
OX NCBI_TaxID=180856;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=97-27;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RT "Complete genome sequence of Bacillus thuringiensis 97-27.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017355; AAT63966.1; -; Genomic DNA.
DR GO; GO:0009986; C:cell surface; IEA.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003885; LRR_cyst.

```

---

```

DR InterPro; IPR007092; LRR_SDS22.
DR InterPro; IPR006635; NEA_transpt.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF00560; LRR_1; 8.
DR Pfam; PF05031; NEAT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00365; LRR_SD22; 8.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50978; NEAT; 1.
KW Complete proteome.
SQ SEQUENCE 954 AA; 108589 MW; 4F4CF8B44C9B355F CRC64;

Query Match 15.6%; Score 101.5; DB 2; Length 954;
Best Local Similarity 25.7%; Pred. No. 26;
Matches 43; Conservative 32; Mismatches 43; Indels 49; Gaps 9;

Qy 1 EDFILPVY--KGEL-----EKGYQF--DGWEIS-GFEG--- 28
Db 727 EIQVPVYDLEGEIENIKLTSEDGTFNNGVKWSTPGKVKYKFDLSDSEISFNGTVI 786

Qy 29 -----KKDAGVIVNLKDTFKVPVKIEEKKKEENKPTDVSKKDNQPNVHSHLN 80
Db 787 QNIVEKEEKEPTKEVEEKEEKEPT-KEVEEKEEKEEPTKEVEEKEEKEEPTKEVEE 845

Qy 81 ESHRKEDLQREHSOKSDSTQVATVLDKN-----NISSKSTNNPN 123
Db 846 ES--KEEV--KEPTKEVEEKEEVAQEIEKSKBEINQSAFVQEQNVN 888

RESULT 23
Q50LX8 ENTHI
ID Q50LX8 ENTHI PRELIMINARY; PRT; 296 AA.
AC Q50LX8
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORENAMES=657.t00001;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RA PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alsmark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich E., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoeft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitch E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.B., Bhattacharya S., Lohia A.,
RA Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFB01001439; BAL42595.1; -; Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 296 AA; 33757 MW; 3A5986BB34A7FC3B CRC64;

Query Match 15.5%; Score 101; DB 2; Length 296;
Best Local Similarity 28.5%; Pred. No. 8;
Matches 35; Conservative 24; Mismatches 44; Indels 20; Gaps 5;

Qy 9 KGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFKVPVKIEEKEE---EENKPTFDV 65
Db 145 EGDSEKKH-----DIPTNEGKENK----DTTKDKNDKEKKDVTWEEGESSKQKTNEE 195

```









```
RC STRAIN=17XXNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABL01000781; EAA22302.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR InterPro; IPR007005; XAP5.
DR PANTHER; PTHR12722; XAP5; 2.
DR Pfam; PF04921; XAP5; 1.
DR SQ SEQUENCE 393 AA; 46652 MW; 55B30519B8FA97D2 CRC64;

Query Match 15.1%; Score 98.5; DB 2; Length 393;
Best Local Similarity 32.7%; Pred. No. 17;
Matches 35; Conservative 14; Mismatches 47; Indels 11; Gaps 5;

QY 26 PEGKKDAGVIVNLSKDT-----PIKVPFKIEKKEENKPTFDVSKKKNPQVNHQ 78
DB 90 FSKEDITY-NESSKDTKHKTNFKLSFFSDDEEEDDEEDKNDENKSETPK-NKSD 147
QY 79 LNESHKEDLQREHSQKSDTKDVTATVLDKNNISKS-TTNNPNK 124
DB 148 EN-SLEKQNEKEBAKESNETEQINKYTKNLQNGKSVNTENK 193

RESULT 33
Q7RL7 PLAYO
ID Q7RL7_PLAYO PRELIMINARY; PRT; 674 AA.
AC Q7RL7;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE Notchless-related.
GN Name=PY02598;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=17XXNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow S.J., van Aken S.B., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABL01000781; EAA22065.1; -; Genomic_DNA.
DR HSSP; P16649; 1ERJ.
DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 8.
DR PRINTS; PRO0320; GPROTEINBRPT.
DR PRODom; PD000018; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS00082; WD_REPEATS_2; 6.
DR PROSITE; PS02094; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 674 AA; 76535 MW; E7521B469FE8E0F7 CRC64;

Query Match 15.1%; Score 98.5; DB 2; Length 674;
Best Local Similarity 24.4%; Pred. No. 30;
Matches 30; Conservative 22; Mismatches 44; Indels 27; Gaps 5;

QY 5 LPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSKDTFIK-----PVFKIEE--- 53
DB 524 LAVFRGHVGPAYKI-AWSI-----DNNYIISQSDSTLKLWINHLVPLKKKEBNGD 575
QY 54 -----KKEENKPTFDV--SKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVT 105
DB 576 QPKSDPKNDQKNDQKDDQKDDQKDDQKDDQKDDQKDDQKDDQKDDQKDDQKDDQK 635
QY 106 TVL 108
DB 636 TLL 638

RESULT 34
Q55G46 DICDI
ID Q55G46_DICDI PRELIMINARY; PRT; 827 AA.
AC Q55G46;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=DD80189573;
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivo F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhoun A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Lousaged H., Mungall K., Oliver K., Price C., Quail M.A., Sanders M.,
RA Urushihara H., Hernandez J., Rabinowitch E., Steffen D., Tivey A.,
MA J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shaulesky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0:0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAFI01000009; EAL73367.1; -; Genomic_DNA.
KW Hypothetical protein; Nucleotide-binding.
SQ SEQUENCE 827 AA; 92700 MW; D576462666B4469 CRC64;

Query Match 15.1%; Score 98.5; DB 2; Length 827;
Best Local Similarity 26.9%; Pred. No. 38;
Matches 29; Conservative 17; Mismatches 31; Indels 31; Gaps 4;

QY 47 VFKKIEBK--EENKPTFDVSKKKNPQVNHSQLNE--SHRKEDLQREH----- 93
```

```
Db 359 VIKLEERKQKEENSNNFGSKRRKSKYNSESESDSRSDIDFDDHNEHKSKI 418
QY 94 -----SQKSDSTKDV-----ATVLDKNNISSKSTNNPN 123
Db 419 KNISKEKKNLSNEHSDNEKDVNRNKIQTQLLDNNNNNNNNKNNNN 466

RESULT 35
OS 096229 PLAF7 PRELIMINARY; PRT; 951 AA.
AC 096229;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PFB0680W.
GN Name=PFB0680W;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99021743; PubMed=9804551; DOI=10.1126/science.282.5391.1126;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aton C., Lai Z., Schwartz D.C., Pertea M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RA "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AF001410; AAC71925.2; -; Genomic_DNA.
DR PIR; B71609; B71609.
KW Hypothetical protein.
SQ SEQUENCE 951 AA; 112486 MW; AC8D889358A84F4P CRC64;

Query Match 15.1%; Score 98; DB 2; Length 951;
Best Local Similarity 27.1%; Pred. No. 48;
Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;

QY 13 EKGQFDGWEI--SFEGKKDAGYVNLKDTFIKVPVKKEEKE-----EENK 60
Db 159 EKGQ-----DISNSAENKCD-----VKEGVKEEKKKEKESIDHKKVEENK 202
QY 61 PTFFD----VSKKKDNPQVNHSQLNSHRKEDLQR-EEHSQKSDSTKQVDTATVLDKNNISS 115
Db 203 KSDHKKVEENKSDHKKVEENKSDHKKIEVKVKEEHEDEE-----DKKEKS 253

QY 116 KSTNNPNK 124
Db 254 ENKNDENK 262

RESULT 36
Q95S93 DROME
ID Q95S93 DROME PRELIMINARY; PRT; 556 AA.
AC Q95S93;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
```

```
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DB GM05229P.
GN Name=Aeph; ORFName=CG8421;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celiker S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY060905; AAL28453.1; -; mRNA.
DR FlyBase; FBgn0034075; Aeph.
DR FlyBase; FBgn0034075; CG8421.
SQ SEQUENCE 556 AA; 63089 MW; 95D32EAC57D11FE8 CRC64;

Query Match 15.0%; Score 97.5; DB 2; Length 556;
Best Local Similarity 24.5%; Pred. No. 29;
Matches 34; Conservative 28; Mismatches 52; Indels 25; Gaps 5;

QY 1 EDFILPVYKGELEKGYQFGM-----EISGFGKKDAGYVI-----NLSKDTFIK 45
Db 78 EDLDTPLSESRFSK--VFDGWDVDEHGDHGVQEPGSGALDDHDDHDDHDEDEE 135
QY 46 PVFKKIEEKEENKPT-----FDVSKKKDNPQVNHSQLNSHRKEDLQREHSQKSDS 99
Db 136 PLTEELEEELEEEETPEDEPAADDEYDEDEENNA--GENITAEADAVEEEEDND 193

QY 100 TKQVDTATVLDKNNISSKT 118
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 37
TH00 YEAST
ID TH00 YEAST STANDARD; PRT; 1345 AA.
AC P38800;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAY-2005 (Rel. 47, Last annotation update)
DE Hypothetical 149.7 kDa protein in IRE1-KSPI intergenic region.
GN OrderedLocusName=YHR080C;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=S288C / AB972;
RX MEDLINE=94376003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favellio A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevaekis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RL VIII.";
RL Science 265:2077-2082(1994).
CC -I- SIMILARITY: Contains 1 GRAM domain.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
```





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:50:52 ; Search time 13.038 Seconds  
(without alignments)  
915.083 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773  
Perfect score: 651  
Sequence: 1 EDFILPVYKGELEKGYQFDG.....ATVLDKNNISSKSTNNPNK 124

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	100.0	2140	2 F95074	serine proteinase,
2	648	99.5	2144	2 A97942	metalloproteinase,
3	108	16.6	558	2 T18467	hypothetical prote
4	99.5	15.3	1038	2 J54977	claustrin - chicke
5	99	15.2	211	2 T25911	hypothetical prote
6	98	15.1	665	2 B71609	hypothetical prote
7	97.5	15.0	1345	2 S46817	hypothetical prote
8	95	14.6	348	2 I37271	cyclicin II - human
9	94	14.4	210	2 T28771	hypothetical prote
10	93	14.3	535	2 T37189	hypothetical prote
11	92.5	14.2	219	2 B72291	hypothetical prote
12	92.5	14.2	312	2 G81339	probable membrane
13	91	14.0	253	2 T32879	hypothetical prote
14	90	13.8	325	2 T18283	hypothetical prote
15	89.5	13.7	508	2 B81594	hypothetical prote
16	89.5	13.7	508	2 C72074	hypothetical prote
17	89.5	13.7	508	2 C72074	hypothetical prote
18	89	13.7	528	2 B96795	unknown protein F2
19	89	13.7	622	2 A90570	lipoprotein (impor
20	89	13.7	1397	2 T10466	DNA topoisomerase
21	88.5	13.6	301	2 T33068	hypothetical prote
22	88.5	13.6	385	2 T20410	hypothetical prote
23	88	13.5	276	2 T23451	hypothetical prote
24	88	13.5	700	2 S67610	probable membrane
25	87.5	13.4	644	2 T47835	hypothetical prote
26	87.5	13.4	1702	2 A1859	Iga-specific metal
27	87.5	13.4	2523	2 T18477	hypothetical prote
28	87	13.4	635	2 F71621	hypothetical prote
29	87	13.4	2464	1 QRMSP1	microtubule-associ

30	86.5	13.3	762	2 G88436	protein T04A8.13 [
31	86.5	13.3	791	2 T24435	hypothetical prote
32	86.5	13.3	2485	1 H71621	serine/threonine-s
33	86	13.2	1016	2 T19006	ankyrin related pr
34	85	13.1	335	2 T33457	hypothetical prote
35	85	13.1	1165	2 H71609	hypothetical prote
36	85	13.1	1202	1 S05362	probable DNA-dirc
37	85	13.1	1976	2 T56555	sodium channel pro
38	85	13.1	3006	2 T28625	variant-specific s
39	84.5	13.0	264	2 A44969	circumsporozoite p
40	84.5	13.0	645	2 B89883	conserved hypotbet
41	84.5	13.0	1694	2 H64106	Iga-specific metal
42	84	12.9	312	2 T25994	hypothetical prote
43	84	12.9	417	2 T49847	hypothetical prote
44	83.5	12.8	143	2 T29740	hypothetical prote
45	83.5	12.8	281	2 T29150	hypothetical prote

ALIGNMENTS

RESULT 1

F95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4  
C;Species: Streptococcus pneumoniae  
C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C;Accession: F95074  
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei  
on, J.D.; Mayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison  
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A;Reference number: A95000; MUID:21357209; PMID:11463916  
A;Accession: F95074  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2140 <KUR>  
A;Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1  
A;Experimental source: strain TIGR4  
C;Genetics:  
A;Gene: SP0641

Query Match	100.0%;	Score	651;	DB	2;	Length	2140;
Best Local Similarity	100.0%;	Pred. No.	7.5e-44;				
Matches	124;	Conservative	0;	Mismatches	0;	Indels	0;
						Gaps	0;
Qy	1	EDFILPVYKGELEKGYQFDGWEISGFGKDGAGVYINLSKDTPIKPVFKIEKKBEENK	60				
Db	1983	EDFILPVYKGELEKGYQFDGWEISGFGKDGAGVYINLSKDTPIKPVFKIEKKBEENK	2042				
Qy	61	PTPDVSKKDNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN	120				
Db	2043	PTPDVSKKDNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN	2102				
Qy	121	NPVK 124					
Db	2103	NPVK 2106					

RESULT 2

A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C;Species: Streptococcus pneumoniae  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C;Accession: A97942  
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burtgett, S.; DeHoff, B.S.;  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.  
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234













QY 109 DK 110  
DB 370 EK 371

## RESULT 23

T23451  
hypothetical protein K08E3.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T23451  
R:McMurray, A.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19743  
A:Accession: T23451  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-276 <WIL>  
A:Cross-references: UNIPROT:Q9XUT1; UNIPARC:UPI0000077D88; EMBL:Z81568; PIDN:CAB04590.1;  
A:Experimental source: clone K08E3  
C:Genetics:  
A:Gene: CESP:K08E3.2  
A:Map position: 3  
A:Introns: 81/3; 102/3; 169/1; 211/2

Query Match 13.5%; Score 88; DB 2; Length 276;  
Best Local Similarity 31.5%; Pred. No. 5.8;  
Matches 28; Conservative 10; Mismatches 29; Indels 22; Gaps 3;

QY 38 LSKDTFIRPVFKKIEKKKEE--NKPTDVSKKKNPQVNHSQLNESHKEDLQREHSQ 95  
DB 60 LEKDEVEKKKKEGKEEKEGKKEKKDKKKD-----SKEDKKDEDHEK 107

QY 96 KSDSTKDVATVLDKNNISSKSTNNPNK 124  
DB 108 KK-----FAEKENNEKKDKNK 128

## RESULT 24

S67610  
probable membrane protein YDL074c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D2483  
C:Species: Saccharomyces cerevisiae  
C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: S67610  
R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67608  
A:Accession: S67610  
A:Molecule type: DNA  
A:Residues: 1-700 <WAM>  
A:Cross-references: UNIPROT:Q07457; UNIPARC:UPI0000069EFF; EMBL:Z74122; NID:G1431087; PI  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:BRE1; MIPS:YDL074c  
A:Cross-references: UNIPROT:Q07457; UNIPARC:UPI0000069EFF; EMBL:Z74122; NID:G1431087; PI  
A:Map position: 4L  
C:Keywords: transmembrane protein  
F:69-85/Domain: transmembrane #status predicted <TMM>

Query Match 13.5%; Score 88; DB 2; Length 700;  
Best Local Similarity 23.7%; Pred. No. 16;  
Matches 31; Conservative 24; Mismatches 40; Indels 36; Gaps 5;

QY 11 ELEKGYQDGEWISGFEKK-----DAGVIVNL-SKDTF---IK 45  
DB 447 DLEKGF-----ELSDLTHKKYSBIINHESVISKLTVEKTKADQKYFAAMRSKDSILIEIK 502

QY 46 PVFKKIEKKKEENKPTDVSKKKNPQVNHSQLNESHKEDLQREHSQSDSKDVTA 105  
DB 503 TLKSKLSKSNEL-----ILQKSDRLLOKQIGNLHKQLDLSQNNERLIDSKTYTL 555

QY 106 TVLDKNNISSK 116

DB 556 KIIDLNTSTK 566

## RESULT 25

T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004  
C:Accession: T47835  
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24475  
A:Accession: T47835  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <NVA>  
A:Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI000009E0AD; EMBL:AL1138658  
A:Experimental source: cultivar Columbia; BAC clone T209  
C:Genetics:  
A:Map position: 3  
A:Introns: 158/2; 329/3  
A:Note: T209.90

Query Match 13.4%; Score 87.5; DB 2; Length 644;  
Best Local Similarity 24.4%; Pred. No. 16;  
Matches 31; Conservative 18; Mismatches 51; Indels 27; Gaps 3;

QY 2 DFILPVVKGLEKGYQDGEWISGFEKKGDAGYVIVNLKDTFIKPVFKKIEKKKEENKP 61  
DB 537 DFLKRIKKNSPQK-----ETTSKNQKNDGNV-----KKENDHQK 573

QY 62 TFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDS-----TKDVTATVLDKNNISSK 117  
DB 574 KSDGNVKKENSKVPRELRSSTGKKVVENNNSKSKRKQTKETAETAEVATGKRGESGK 633

QY 118 TTNPNK 124  
DB 634 DDKQPRK 640

## RESULT 26

A41859  
IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae  
C:Species: Haemophilus influenzae  
A:Variety: strain HK715  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: A41859  
R:Poulsen, K.; Reinholdt, J.; Killian, M.  
J. Bacteriol. 174, 2913-2921, 1992  
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae  
A:Reference number: A41859; MUID:92234949; PMID:1373717  
A:Accession: A41859  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1702 <POU>  
A:Cross-references: UNIPROT:P45384; UNIPARC:UPI000012D3F0; GB:M87489; NID:G148906; PIDN:  
A:Experimental source: strain HK715  
A:Note: sequence extracted from NCBI backbone (NCBIP:97282)  
C:Superfamily: IGA-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match 13.4%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 49;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 36 INLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHS- 94  
DB 1296 INTGSATAITETAESKDKPQTETAATEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY 95 --QKSDSTKDVATVLDKNNISSKSTNNPNK 124

DB 1356 SQPQTSABETTAASTDETTIADNSKSKPNR 1387



[illegible]

RESULT 31

T24435  
hypothetical protein T04A8.13 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T24435  
R:Palmer, S.  
submitted to the EMBL Data Library, August 1994  
A:Reference number: Z19889  
A:Accession: T24435  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-791 <WIL>  
A:Cross-references: UNIPROT:Q22142; UNIPARC:UPI00000061132; EMBL:Z35663; PIDN:CAA84732.2;  
A:Experimental source: clone T04A8  
C:Genetics:  
A:Gene: CESP:T04A8.13  
A:Map position: 3  
A:Introns: 31/3; 212/1; 229/3; 331/3; 406/1; 472/3; 572/1; 651/2

**RESULT 32**

H71621  
serine/threonine-specific protein kinase (EC 2.7.1.-) PF80150c - malaria parasite (Plasmodium falciparum)  
C/Species: Plasmodium falciparum  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Dec-2004  
C/Accession: H71621  
R; Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; R; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1126-1132, 1998  
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A/Reference number: A71600; MUID:99021743; PMID:9804551  
A/Accession: H71621  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-2485 <GAR>  
A/Cross-references: UNIPROT:O96134; UNIPARC:UPI0000172553; GB:AE001362; NID  
A/Experimental source: clone 3D7  
C/Genetics:  
A/Gene: PF80150c  
C/Superfamily: protein kinase homology  
C/Keywords: phosphotransferase  
F/2087-2352/Domain: protein kinase homology <KIN>

**Q7** 94 ---SQSDSTKDTATVLDKNNISSKSTTN 121  
|:::|::|||:::|:  
**D6** 2080 INBEHKNEGKLTYYNWKNNISNNYND 2110

RESULT 33

T19006  
 ankryrin related protein C06C3.1 - *Caenorhabditis elegans*  
 N;Contains: myosin-light-chain-phosphatase (EC 3.1.3.53)  
 C;Species: *Caenorhabditis elegans*  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C;Accession: T19006; T22086  
 R;Berks, M.  
 submitted to the EMBL Data Library, August 1994  
 A;Reference number: Z19058  
 A;Accession: T19006  
 A;Status: preliminary; translated from GB/EMBL/DBBJ  
 A;Molecule type: DNA  
 A;Residues: 1-1016 <W1>  
 A;Cross-references: UNIPROT:Q17718; UNIPARC:UPI0000081D09; EMBL:Z36719; PIDN:CAAB5318.  
 A;Experimental source: clone C06C3  
 R;Matthews, P.  
 submitted to the EMBL Data Library, January 1995  
 A;Reference number: Z19510  
 A;Accession: T22086  
 A;Status: preliminary; translated from GB/EMBL/DBBJ  
 A;Molecule type: DNA  
 A;Residues: 1-1016 <W1>  
 A;Cross-references: UNIPARC:UPI0000081D09; EMBL:Z47809; PIDN:CAAB7782.1; GSPDB:GN000020.0  
 A;Experimental source: clone F42A8  
 C;Genetics:  
 A;Gene: CRSP:C06C3.1  
 A;Map position: 2  
 A;Introns: 27/3; 94/3; 279/3; 352/2; 462/2; 523/3; 569/2; 657/2; 718/3; 766/3; 833/3;  
 C;Keywords: phosphoric monoester hydrolase

RESULT 34

T33457  
hypoetical protein F36H12.3 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 23-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T33457  
R/Blanchard, M.; Bradshaw, H.; Stellyes, L.  
submitted to the EMBL Data Library, July 1998  
A/Description: The sequence of C. elegans cosmid F36H12.  
A/Reference number: Z21346  
A/Accession: T33457  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-335 <BLA>  
A/Cross-references: UNIPROT:076719; UNIPARC:UP10000076COA; EMBL:AF078790; P  
A/Experimental source: strain Bristol N2; clone F36H12  
C/Genetics:  
A/Gene: CESP:F36H12.3  
A/Map position: 4

A;Introns: 32/3; 227/1; 270/2

Query Match 13.1%; Score 85; DB 2; Length 335;  
Best Local Similarity 27.3%; Pred. No. 12;  
Matches 30; Conservative 12; Mismatches 46; Indels 22; Gaps 2;

Qy 15 GYQFDGWEISGPEG-KKDAGYVIN-----LSKDTFTKPVFKKIE 52

Db 10 GITVAGWILVGGGKKKAGAVKSTAPAPAKADSKMKAPBEKSKSEKKEPKKEE 69

Qy 53 EKKEBEKPTFDVSKKKNQPNVHSQLNESHKEDLQREHSQKSDSTKD 102

Db 70 EKKEKSKSEKSKDKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEK 119

RESULT 35  
H71609  
hypothetical protein PFB0630c - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C;Accession: H71609  
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A;Reference number: A71600; MUID:99021743; PMID:9804551  
A;Accession: H71609  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1166 <GAR>  
A;Cross-references: UNIPROT:O96219; UNIPARC:UPI000007D874; GB:AE001407; GB:AE001362; NID  
A;Experimental source: clone 3D7  
C;Genetics:  
A;Gene: PFB0630c

Query Match 13.1%; Score 85; DB 2; Length 1166;  
Best Local Similarity 28.6%; Pred. No. 50;  
Matches 28; Conservative 16; Mismatches 34; Indels 20; Gaps 3;

Qy 28 GKGDAGYVNLKSDTFIKPVFKKIEKBEENKPTFDVSKKKNQPNVHSQLNESHKED 87

Db 614 GKKEE-----KCVTKIIEIKNEEMEPNQEQDKDN-----ELKEKNDKEE 655

Qy 88 L--QREHSQKSDSTKQVTVTLVDKNNISSKSTNNPN 123

Db 656 KNDQEKNDQBEKNDKEKNDKEKNDQEKNDQEKNDQEKND 693

RESULT 36  
S05362  
probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascobolus immersus) mitochondrion  
C;Species: mitochondrion Ascobolus immersus  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Dec-2004  
C;Accession: S05362  
R;Kempken, F.; Meinhardt, F.; Esser, K.  
Mol. Gen. Genet. 218, 523-530, 1989  
A;Title: In organello replication and viral affinity of linear, extrachromosomal DNA of  
A;Reference number: S05362; MUID:90066356; PMID:2573821  
A;Accession: S05362  
A;Molecule type: DNA  
A;Residues: 1-1202 <KEM>  
A;Cross-references: UNIPROT:P22374; UNIPARC:UPI0000129831; EMBL:X15982; NID:g2933; PIDN:

Query Match 13.1%; Score 85; DB 1; Length 1202;  
Best Local Similarity 25.6%; Pred. No. 52;  
Matches 33; Conservative 21; Mismatches 55; Indels 20; Gaps 5;

Qy 9 KGELEKGYQFDGWEISGPEGKGDAGYVNLKSDTFIKPVFKKIE-----EKKEBEK-P 61

Db 372 KGSL----SPDFKPLKTIEGTYANVTFFPKKDIIWVKDKINKKINFNGLDLPKTMDLSKWP 427

Qy 62 TFDVSKKKNQPNVHSQLNESHKEDLQREHSQKSDSTKQVTVTLVDKNNI-----S 114

Db 428 NLKLNKORTSGEIRMTTKNNKNSQSYDI--IGHMINDGENVITFNRAVDNSIIKFTVTD 485

Qy 115 SKSTNNPN 123

Db 486 SMGVTNDPN 494

RESULT 37  
I56555  
sodium channel protein 6 - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 09-Jul-2004  
C;Accession: I56555  
R;Schaller, K.L.; Krzemien, D.M.; Yarowsky, P.J.; Krueger, B.K.; Caldwell, J.H.  
J. Neurosci. 15, 3231-3242, 1995  
A;Title: A novel, abundant sodium channel expressed in neurons and glia.  
A;Reference number: I56555; MUID:95271234; PMID:7751906  
A;Accession: I56555  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1976 <RES>  
A;Cross-references: UNIPROT:Q63541; UNIPARC:UPI00000E7CC6; GB:L39018; NID:g829033; PIDN:  
C;Genetics:  
A;Gene: SCP6  
C;Superfamily: sodium channel protein  
C;Keywords: duplication

Query Match 13.1%; Score 85; DB 2; Length 1976;  
Best Local Similarity 25.2%; Pred. No. 91;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

Qy 17 QFDGWEISGPEGKGDAGYVNLKSDTFIKPVFKKIEKBEENKPTFDVSKKKNQPNVH 76

Db 989 EMNNLQISVIRIKGVAVT-KVKVHAFMQAHFK--QREADEVKPLDELYEKKANCIAH 1044

Qy 77 SQLNESHKEDLQREHSQKSDSTKQVTVTLVDKNNISSKSTNNPN 123

Db 1045 TGV-DIHRNGDFKNGTGTSGIGSVSEKVIIDEDHM---SFINNPN 1087

RESULT 38  
T28625  
variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T28625  
R;Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, D.S.  
Cell 82, 89-100, 1995  
A;Title: The large diverse gene family var encodes proteins involved in cytoadherence an  
A;Reference number: Z20487; MUID:95330813; PMID:7606788  
A;Accession: T28625  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-3006 <SUX>  
A;Cross-references: UNIPROT:Q26032; UNIPARC:UPI000007B568; EMBL:L40609; NID:g886376; PID  
C;Genetics:  
A;Gene: var-3  
A;Introns: 2597/3

Query Match 13.1%; Score 85; DB 2; Length 3006;  
Best Local Similarity 27.3%; Pred. No. 1.5e+02;  
Matches 35; Conservative 20; Mismatches 53; Indels 20; Gaps 5;

Qy 9 KGELEK-----GYQFDGWEIS-GFEGKKDAGYVNLK-----DTFTKPVFKKIEE 53

Db 2134 KGATEKKQKDAEGNNNDYKEFSKTYRNVDAAFLNLSLNGPCSKNDDSDVQDEIKFDE 2193

Qy 54 KK-----EENKPTFDVSKK-KDNPCVNHSQNLSESHRKEDLQREHSQKSDSTKQVTVTL 108



Job time : 16.038 secs

Db 2194 RKTFGHTYCKPSKITVCKENNHCDNSKPNDCRNINSISAEDIEKRSNTQDVTMSVS 2253

QY 109 DKNNTSSK 116

Db 2254 DSNTNGNK 2261

## RESULT 39

A44969

circumsporozoite protein precursor - Plasmodium yoelii nigeriensis

C:Species: Plasmodium yoelii nigeriensis

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 01-Dec-1995

C:Accession: A44969

R:Colomer-Gould, V.; Enea, V.

Mol. Biochem. Parasitol. 43, 51-58, 1990

A:Title: Plasmodium yoelii nigeriensis circumsporozoite gene structure and its implicati

A:Reference number: A44969; MUID:91148645; PMID:2290446

A:Accession: A44969

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-264 &lt;COL&gt;

A:Cross-references: UNIPARC:UPI000177F73; GB:M32350

C:Superfamily: circumsporozoite protein; thrombospondin type 1 repeat homology

F:190-242/Domain: thrombospondin type 1 repeat homology &lt;THR1&gt;

Query Match 13.0%; Score 84.5; DB 2; Length 264;

Best Local Similarity 28.2%; Pred. No. 10;

Matches 29; Conservative 14; Mismatches 29; Indels 31; Gaps 5;

QY 45 KVPFKK-----IEKKK-----EENKPTFDVSKKDNQVNHSQLNESHKEDLQREE----- 92

Db 75 KPAKKDLPKBEKKDDLPKBEKKDDLPKBEKKDDPP-----KEKKDDLPKEEKDA 127

QY 93 -----HSQKSDSTKDTATVL--DKNNISSKSTNNPN 123

Db 128 PFGNKDAPKEKKADPPKEAQNKLPQDNNNNNNNNNNNN 170

## RESULT 40

E89883

conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (strain N315)

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C:Accession: E89883

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayaashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: E89883

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-645 &lt;KUR&gt;

A:Cross-references: UNIPROT:Q99UK5; UNIPARC:UPI00000CAE3; GB:BA000018; PID:gl3700929; E

A:Experimental source: strain N315

C:Genetics:

A:Gene: SA0976

Query Match 13.0%; Score 84.5; DB 2; Length 645;

Best Local Similarity 28.6%; Pred. No. 28;

Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;

QY 31 DAGYVNL-SKDTETKPVFKIEKKKEENKPTFDV-----SKKDNQVNHSQLNESH 84

Db 441 DQYHVRIVDKAFTKANTDKNKKEQQDNSAKKATPATPSKPTSPVKEKQDSOK 500

QY 85 KEDLQ----REHSOKSDSTKQVT--ATVLDKNNISSKSTNNPNK 124

Db 501 DDNKQLPSVEKENDASSEGGKDKTPATKPTGEVSSSTT--PTK 543

Search completed: April 24, 2006, 15:01:11

**THIS PAGE BLANK (user)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 9.70917 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773

Perfect score: 651

Sequence: 1 EDFILPVYKGELEKGYQFDG.....ATVLDKNNISKSTTNNPK 124

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:\*

- 1: /SIDSS/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /SIDSS/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /SIDSS/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /SIDSS/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /SIDSS/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 6: /SIDSS/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /SIDSS/prodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /SIDSS/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	16.3	354	7 US-11-189-817-2	Sequence 2, Appli
2	101	15.5	651	7 US-11-128-660-1	Sequence 1, Appli
3	86	13.2	443	7 US-11-188-298-1015	Sequence 1015, Ap
4	85	13.1	472	6 US-10-793-626-658	Sequence 658, App
5	85	13.1	720	6 US-10-793-626-2058	Sequence 2058, Ap
6	85	13.1	746	6 US-10-793-626-652	Sequence 652, App
7	84.5	13.0	501	6 US-10-485-517-381	Sequence 381, App
8	84.5	13.0	645	6 US-10-485-517-244	Sequence 244, App
9	84.5	13.0	1694	7 US-11-052-554A-83	Sequence 83, Appl
10	82	12.6	439	7 US-11-188-298-15964	Sequence 15964, A
11	82	12.6	439	7 US-11-188-298-16606	Sequence 16606, A
12	82	12.6	886	7 US-11-087-099-11456	Sequence 11456, A
13	79.5	12.2	140	7 US-11-096-568A-4771	Sequence 4771, Ap
14	79.5	12.2	943	6 US-10-475-204-34	Sequence 34, Appl
15	79.5	12.2	1036	7 US-11-096-568A-28315	Sequence 28315, A
16	79.5	12.2	1070	7 US-11-096-568A-28314	Sequence 28314, A
17	79.5	12.2	1276	7 US-11-096-568A-28313	Sequence 28313, A
18	78.5	12.1	258	7 US-11-098-686-10475	Sequence 10475, A
19	78.5	12.1	700	7 US-11-196-475-74	Sequence 74, Appl
20	78.5	12.1	1155	6 US-10-793-626-1780	Sequence 1780, Ap
21	77.5	11.9	405	6 US-10-485-517-239	Sequence 239, App
22	77.5	11.9	627	6 US-10-873-528-191	Sequence 191, App
23	77	11.8	663	7 US-11-196-475-70	Sequence 70, Appl
24	77	11.8	663	7 US-11-196-475-70	Sequence 78, Appl
25	76.5	11.8	439	7 US-11-188-298-17915	Sequence 17915, A

26	76.5	11.8	568	6 US-10-793-626-2482	Sequence 2482, Ap
27	76.5	11.8	1145	6 US-10-793-626-1432	Sequence 1432, Ap
28	76	11.7	244	7 US-11-096-568A-11931	Sequence 11931, A
29	76	11.7	482	7 US-11-232-440-33	Sequence 33, Appl
30	75.5	11.6	278	6 US-10-793-626-2700	Sequence 2700, Ap
31	75.5	11.6	404	6 US-10-793-626-398	Sequence 398, App
32	75.5	11.6	558	6 US-10-510-903-22	Sequence 22, Appl
33	75	11.5	513	7 US-11-185-924-16	Sequence 16, Appl
34	74.5	11.4	313	7 US-11-096-568A-32043	Sequence 32043, A
35	74.5	11.4	425	7 US-11-096-568A-32042	Sequence 32042, A
36	74.5	11.4	434	7 US-11-096-568A-32041	Sequence 32041, A
37	74.5	11.4	693	7 US-11-196-475-68	Sequence 68, Appl
38	74	11.4	550	7 US-11-087-099-5463	Sequence 5463, Ap
39	74	11.4	1758	7 US-11-087-099-9570	Sequence 9570, Ap
40	73.5	11.3	417	6 US-10-821-234-1536	Sequence 1536, Ap
41	73.5	11.3	493	7 US-11-096-568A-3070	Sequence 3070, Ap
42	73.5	11.3	493	7 US-11-096-568A-3071	Sequence 3071, Ap
43	73.5	11.3	510	7 US-11-096-568A-3069	Sequence 3069, Ap
44	73	11.2	472	6 US-10-793-626-1902	Sequence 1902, Ap
45	73	11.2	603	7 US-11-024-959-484	Sequence 484, App

ALIGNMENTS

RESULT 1  
US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US2006030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRUIHHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM  
; TITLE OF INVENTION: MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES  
; FILE REFERENCE: 275601US0  
; CURRENT APPLICATION NUMBER: US/11/189,817  
; CURRENT FILING DATE: 2005-07-27  
; PRIOR APPLICATION NUMBER: 60/598,062  
; PRIOR FILING DATE: 2004-08-03  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2  
; LENGTH: 354  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-11-189-817-2

Query Match 16.3%; Score 106; DB 7; Length 354;  
Best Local Similarity 23.5%; Pred. No. 0.046;  
Matches 32; Conservative 29; Mismatches 49; Indels 26; Gaps 5;

Qy	7	VYKGELEKGYQFD-GWEISGF--EGKKDAG-----YVINLSKDTFIKPVFKKIEKKEE- 57
Db	181	VLKAKASSYDIILGWEFGGVPPEKKKENMLSHLYVSSKDKENISKENDVDLVEKEEA 240
Qy	58	-----ENKPTFDVSKKDNQVNHSQLNESHREKDLQREHSQKSDSTKDVTA 105
Db	241	EETEEELKKNBEETSEISEDEEBEKEEENDKKKEQKESNNENNDQKDMEA 300
Qy	106	TVLDKNNISKSSTNN 121
Db	301	-----QNLISKNNNN 311

RESULT 2  
US-11-128-660-1  
; Sequence 1, Application US/11128660  
; Publication No. US2006002432A1  
; GENERAL INFORMATION:  
; APPLICANT: Statens Serum Institut  
; TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plasmodium falciparum  
; TITLE OF INVENTION: falciparum  
; FILE REFERENCE: 15007dk

```
; CURRENT APPLICATION NUMBER: US/11/128,660
; CURRENT FILING DATE: 2005-05-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-128-660-1

Query Match      15.5%; Score 101; DB 7; Length 651;
Best Local Similarity 25.4%; Pred. No. 0.25;
Matches 31; Conservative 26; Mismatches 39; Indels 26; Gaps 5;

Qy 20 GWEISGF--EGKDGAG-----VIVNLSKDTFTKPVFKIEEKKKEENKPTFDVSKKDNPP 72
Db 493 GWFGGVPVHKKEENMLSHLYSSKOKENISKENDVDLDE--KEEABETBEELDEEKN 551

Qy 73 QVNHSQLN-----ESHKEDLQREHSQKSDSTKDVTVATVLDKNNISSKSTT 119
Db 552 BETESEISDEEEEEKEEENKKEQEKQSENNDQKDMA-----QNLISKQN 606

Qy 120 NN 121
Db 607 NN 608

RESULT 3
US-11-188-298-1015
; Sequence 1015, Application US/11/188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452) B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 1015
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Thermococcus kodakaraensis
US-11-188-298-1015

Query Match      13.2%; Score 86; DB 7; Length 443;
Best Local Similarity 17.8%; Pred. No. 3.4;
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

Qy 8 YKGELEKGYQDPGWETSGPEGKKGAGYVINLSKDTFI----- 44
Db 42 YEEAVEVDGVSFGSSIPGPEGIEDSLIFKADPSTYAEIPWEGIGRVGYIYKGDEPYQA 101

Qy 45 --KPVFKIEEKEEN-----KPTFDVSKKD----- 70
Db 102 DPGILKRVLERLEKEGLKAHIGPPEFVIFKNGTWELHIFDPSGGYFDLVGLDKARIR 161

Qy 71 -----NPQVNHSQLNESHRKEDLQREHSQKSD-----STKDVTVATVLD 109
Db 162 REIALYMPYLGLKPEVLHHEVGKAQHEIDFRYDEALRTADNIVSPKHVVKVAE 215

RESULT 4
US-10-793-626-658
; Sequence 658, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258

; CURRENT APPLICATION NUMBER: US/11/128,660
; CURRENT FILING DATE: 2005-05-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-128-660-1

Query Match      15.5%; Score 101; DB 7; Length 651;
Best Local Similarity 25.4%; Pred. No. 0.25;
Matches 31; Conservative 26; Mismatches 39; Indels 26; Gaps 5;

Qy 20 GWEISGF--EGKDGAG-----VIVNLSKDTFTKPVFKIEEKKKEENKPTFDVSKKDNPP 72
Db 493 GWFGGVPVHKKEENMLSHLYSSKOKENISKENDVDLDE--KEEABETBEELDEEKN 551

Qy 73 QVNHSQLN-----ESHKEDLQREHSQKSDSTKDVTVATVLDKNNISSKSTT 119
Db 552 BETESEISDEEEEEKEEENKKEQEKQSENNDQKDMA-----QNLISKQN 606

Qy 120 NN 121
Db 607 NN 608

RESULT 5
US-10-793-626-2058
; Sequence 2058, Application US/10793625
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2058

Query Match      13.1%; Score 85; DB 6; Length 720;
Best Local Similarity 28.0%; Pred. No. 7.2;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

Qy 51 IEKKKEENKPTFDVSKKQD-----NPQVNHSQLNESHRKEDLQREHSQKSDSTK----D 102
Db 38 LEEQIKALDKKFKASQAKQNTQNNHQSNNKQNSNDKKEKQSKNSKPTKCKEQN 97

Qy 103 VTATVLDKNNISSKSTTNNPK 124
Db 98 NKGQKQNNKNTKNNKNNK 119

RESULT 6
US-10-793-626-652
; Sequence 652, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258

; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2005-05-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-128-660-1

Query Match      15.5%; Score 101; DB 7; Length 651;
Best Local Similarity 25.4%; Pred. No. 0.25;
Matches 31; Conservative 26; Mismatches 39; Indels 26; Gaps 5;

Qy 20 GWEISGF--EGKDGAG-----VIVNLSKDTFTKPVFKIEEKKKEENKPTFDVSKKDNPP 72
Db 493 GWFGGVPVHKKEENMLSHLYSSKOKENISKENDVDLDE--KEEABETBEELDEEKN 551

Qy 73 QVNHSQLN-----ESHKEDLQREHSQKSDSTKDVTVATVLDKNNISSKSTT 119
Db 552 BETESEISDEEEEEKEEENKKEQEKQSENNDQKDMA-----QNLISKQN 606

Qy 120 NN 121
Db 607 NN 608

RESULT 5
US-10-793-626-2058
; Sequence 2058, Application US/10793625
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2058

Query Match      13.1%; Score 85; DB 6; Length 720;
Best Local Similarity 28.0%; Pred. No. 7.2;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

Qy 51 IEKKKEENKPTFDVSKKQD-----NPQVNHSQLNESHRKEDLQREHSQKSDSTK----D 102
Db 38 LEEQIKALDKKFKASQAKQNTQNNHQSNNKQNSNDKKEKQSKNSKPTKCKEQN 97

Qy 103 VTATVLDKNNISSKSTTNNPK 124
Db 98 NKGQKQNNKNTKNNKNNK 119

RESULT 6
US-10-793-626-652
; Sequence 652, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
```

; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 652  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-652

Query Match 13.1%; Score 85; DB 6; Length 746;  
Best Local Similarity 26.4%; Pred. No. 7.5;  
Matches 37; Conservative 19; Mismatches 42; Indels 42; Gaps 8;  
QY 4 ILPVVK-----GELEKGVQFDGW---BISGPE-----GKKDAGYVIN--LSKDTPIK 45  
DB 615 VLPHSKVLMWTGELTMP-DMTGWTKEDVLAFEDTLKVKSTKNGFVTVNQSIKGOIILK 673  
QY 46 PVFKKIEKKKEENKPTFDVS-----KKKDNPOVNHSQLNESHKEDLQREHHSQKSDSTK 101  
DB 674 -----NKKIEVLSAEDTDDQEKTDDESDSNKSKKKADEHDSNTSSSTK 720  
QY 102 DVTATVLDKNINSSKSTTNN 121  
DB 721 N-----DKSNADSKNDSDD 734

RESULT 7  
US-10-485-517-381  
; Sequence 381, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, James  
; APPLICANT: Mond, James  
; TITLE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629WO  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; CURRENT FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118825.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 381  
; LENGTH: 501  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-381

Query Match 13.0%; Score 84.5; DB 6; Length 501;  
Best Local Similarity 28.6%; Pred. No. 5.2;  
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;  
QY 31 DAGYVINL-SKDTPIKPVFKKIEKKKEENKPTFDV-----SKKDNPOVNHSQLNESHHR 84  
DB 297 DGQYHVRIVDKCAFTKANTDKSNKKBQDNSAKKEATPATPSKPTSPVKEKESQKQDSOK 356  
QY 85 KEDLQ-----REHHSQKSDSTKQVT--ATVLDKNINSSKSTTNNPK 124  
DB 357 DDNKQLPSVEKENDASSESGKDKTPTATPKYKGEVSSSTT--PTK 399

RESULT 8  
US-10-485-517-244  
; Sequence 244, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield

; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, James  
; APPLICANT: Mond, James  
; TITLE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629WO  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; CURRENT FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118825.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 244  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-244

Query Match 13.0%; Score 84.5; DB 6; Length 645;  
Best Local Similarity 28.6%; Pred. No. 7;  
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;  
QY 31 DAGYVINL-SKDTPIKPVFKKIEKKKEENKPTFDV-----SKKDNPOVNHSQLNESHHR 84  
DB 441 DGQYHVRIVDKCAFTKANTDKSNKKBQDNSAKKEATPATPSKPTSPVKEKESQKQDSOK 500  
QY 85 KEDLQ-----REHHSQKSDSTKQVT--ATVLDKNINSSKSTTNNPK 124  
DB 501 DDNKQLPSVEKENDASSESGKDKTPTATPKYKGEVSSSTT--PTK 543

RESULT 9  
US-11-052-554A-83  
; Sequence 83, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DBL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: Patent in version 3.3  
; SEQ ID NO 83  
; LENGTH: 1694  
; TYPE: PRT  
; ORGANISM: Haemophilus influenzae Rd  
US-11-052-554A-83

Query Match 13.0%; Score 84.5; DB 7; Length 1694;  
Best Local Similarity 26.1%; Pred. No. 22;  
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;  
QY 36 INLSKDTPIKPVFKKIEKKKEENKPTFDVSKKDNPOVNHSQLNESHKEDLQREHHS- 94  
DB 1288 INTGSAITAITETAESKDKPTQTETAATEDASQHKANTVADNSVANNSESDPKSRRRSI 1347  
QY 95 --QKSDSTKQVTATVLDKNINSSKSTTNNPK 124  
DB 1348 SQOETSABETTAASTDETTIADNSKSKPNR 1379

RESULT 10  
US-11-188-298-15964  
; Sequence 15964, Application US/11188298  
; Publication No. US20060075522A1  
; GENERAL INFORMATION:



```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: PROTEIN INTERACTIONS IN VERTEBRATE CELLS
; FILE REFERENCE: HMV-056.25
; CURRENT APPLICATION NUMBER: US/10/475,204
; CURRENT FILING DATE: 2003-10-17
; PRIOR APPLICATION NUMBER: PCT/US02/13008
; PRIOR FILING DATE: 2002-09-27
; PRIOR APPLICATION NUMBER: 60/285,509
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-475-204-34

```

Query Match	12.2%	Score 79.5;	DB 6;	Length 943;
Best Local Similarity	23.1%	Pred. No. 30;		
Matches	36;	Conservative 16;	Mismatches 55;	Indels 49; Gaps 5
Qy	6	PVYKGELEKGQVPDWEISGPEGKDGAVTNLSKOTFIKVPFKIKBKK-----EEN	59	
Db	375	PSDTVLTSTVALDETWNTRSTKYEMSKNAEKPSSKRTIKQQRKFKWAPAESQ-	433	
Qy	60	KPTFDVSKKGD--NPOVNHSQLNSHRKEDLOREH-----SOKSDSTK	101	
Db	434	---LDVGSGKDNIHTSHITQDEPQNSDRNMEHEBNGDCVSKKQMPVPGSKKLSSTK	490	
Qy	102	D-----VTATVLDKKNISK	116	
Db	491	DKESSKKRFSSESKNLVPEVSTSTVTKSRISRR	526	

```

RESULT 15
US-11-096-568A-28315
; Sequence 28315, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28315
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1036)
; OTHER INFORMATION: Ceres Seq. ID no. 2712010
US-11-096-568A-28315

```

```
Query Match      12.2% Score 79.5; DB 7; Length 1036;
Best Local Similarity 23.2%; Pred.No.33;
Matches 23; Conservative 20; Mismatches 47; Indels 9; Gaps 2
```

Qy	27	EGKDAAGVIVNLSKDT?-----IKPVFKJIEEKKEENKPTFDVSKKKONPQNHSOLN	80
		: :	
Db	552	DGDSDKKVMVEVGKKS DSGS VEMKP TAES LEDVKDENAS KTV DVKG TSPDT KKGEA	611
Qy	81	ESHREKDLOREEH---SQKSDSTKDVATVLDKNISSK	116
		: :	
Db	612	SSSSKOOTKTGDEKKAEEKNNSETWSEGGCKIDRNNITDK	650

RESULT 16  
US-11-096-568A-28314  
; Sequence 28314, Application US/11096568A  
; Publication No. US20060048240A1

```

; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide:
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28314
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1070)
; OTHER INFORMATION: Ceres Seq. ID no. 2712009
US-11-096-568A-28314

```

	Query Match	12.2%	Score 79.5;	DB 7;	Length 1070;	
	Best Local Similarity Matches	23.2%;	Pred. No. 35;			
	Matches	23;	Conservative 20;	Mismatches 47;	Indels 9;	Gaps 2;
Qy	27	EGKDAGYVNLNKDTF-----IKPVFKIEBKKEBENKPTDPVSKKKNDPNNISQLN	80			
	:	: :				
Dd	586	DGDSDEKVKMEVGKSGSDGSVEWKPTAELSLEDVKDENASKTVDVQETGSPDTKKKEGA	645			
Qy	81	ESHRKEDLQREKH---SOKSDSTKYDTATVLDKNINSSK	116			
Dd	646	SSSSKKDKTTGGDKAEAKONNSTMGSEKKIDRNNTDEK	684			

```

RESULT 17
US-11-096-568A-28313
; Sequence 28313, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nictolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28313
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1276)
; OTHER INFORMATION: Ceres Seq. ID no. 2712008
US-11-096-568A-28313

```

	Query Match	12.2%	Score 79.5;	DB 7;	Length 1276;
	Best Local Similarity	23.2%;	Pred. No. 43;	Mismatches	23; Conservative 20; Mismatches 47; Indels 9; Gaps 2;
Qy	27 EGKDAGYVNIISKDTF-----IKVPFKIIEKKKEENKPFTFDYVKCKDNFPQNHSQLN	80			
:	: ::  :	::: :: :	:: :	:	:
Dd	792 DGDSDSEKVMEVGKSSDGSSEMKTAEISLEBDVDKENASTVTDVDKETGSPOTTKKEGA	851			
Qy	81 ESHRKEDLQBEEH---SQKSDSTKYDVATVLDDNNISSK	116			
:	: :: :	::: :: :	:: :	:	:
Dd	852 SSSSKDKTTGGDGKAEEKNNSETSGEKKIDRNNITDEK	890			

RESULT is  
US-11-098-686-10475  
; Sequence 10475, Application US/11098686  
; Publication No. US20060024696A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES

```
/ TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
/ FILE REFERENCE: 09531-128001
/ CURRENT APPLICATION NUMBER: US/11/098,686
/ CURRENT FILING DATE: 2005-04-04
/ PRIOR APPLICATION NUMBER: PCT/US03/31318
/ PRIOR FILING DATE: 2003-10-01
/ PRIOR APPLICATION NUMBER: US 60/416,395
/ PRIOR FILING DATE: 2002-10-04
/ NUMBER OF SEQ ID NOS: 11433
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10475
/ LENGTH: 258
/ TYPE: PRT
/ ORGANISM: Lawsonia intracellularis
US-11-098-686-10475

Query Match          12.1%; Score 78.5; DB 7; Length 258;
Best Local Similarity 22.0%; Pred. No. 8.1;
Matches 20; Conservative 26; Mismatches 30; Indels 15; Gaps 3;

QY 38 LSKDTFIKPVFKIEEKEENKPT-FDVSKKK-----DNPQVNHSQLNESHKEDLQ 89
Db 81 IEKQILKPENRUFHMLKQDNPNFTLYNKKKKHNLQSIEDITNLQHSQIPSSHPQDLK 140

QY 90 RE-----EHSQKSDSTKQVTVATVLDKNNI 113
Db 141 OKKDISSETKNQIKNTKEQVQITSQNDL 171

RESULT 19
US-11-196-475-74
/ Sequence 74, Application US/11196475
/ Publication No. US20050271682A1
/ GENERAL INFORMATION:
/ APPLICANT: Dattwyler, Raymond J.
/ APPLICANT: Gomes Solecki, Maria J. C.
/ APPLICANT: Luft, Benjamin J.
/ APPLICANT: Dunn, John J.
/ TITLE OF INVENTION: Recombinant Constructs of Borrelia
/ FILE REFERENCE: 2631.1001-011
/ CURRENT APPLICATION NUMBER: US/11/196,475
/ CURRENT FILING DATE: 2005-08-03
/ PRIOR APPLICATION NUMBER: US 08/148,191
/ PRIOR FILING DATE: 1993-11-01
/ PRIOR APPLICATION NUMBER: US 08/235,836
/ PRIOR FILING DATE: 1994-04-29
/ PRIOR APPLICATION NUMBER: US 09/666,017
/ PRIOR FILING DATE: 2000-09-19
/ PRIOR APPLICATION NUMBER: US 60/226,484
/ PRIOR FILING DATE: 2000-08-18
/ PRIOR APPLICATION NUMBER: PCT/US01/24736
/ PRIOR FILING DATE: 2001-08-07
/ NUMBER OF SEQ ID NOS: 213
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 74
/ LENGTH: 700
/ TYPE: PRT
/ ORGANISM: Borrelia burgdorferi
US-11-196-475-74

Query Match          12.1%; Score 78.5; DB 7; Length 700;
Best Local Similarity 24.3%; Pred. No. 26;
Matches 27; Conservative 29; Mismatches 32; Indels 23; Gaps 5;

QY 36 INLSKDTFIKPVFKIEEKEENKPT-FDVSKKKDNFQV-----HSQLNES----H 83
Db 316 LDIQRTVREKLQENINETNKKLPKPGDVSSPKVDQLQIKESLEDLQELQKETGDN 375

QY 84 RKEDLQREHSQKSD-----STKQVTVATVLDKNNISSKSTNNPNK 124
Db 376 QKREIEKQTEIKSDEKLLKSKDKSKDGKALDLDR-ELNSKASSKSKSK 425
```

## RESULT 20

```
US-10-793-626-1780
/ Sequence 1780, Application US/10793626
/ Publication No. US20050255478A1
/ GENERAL INFORMATION:
/ APPLICANT: KIMMERLY, WILLIAM JOHN
/ TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
/ FILE REFERENCE: PU3480US
/ CURRENT APPLICATION NUMBER: US/10/793,626
/ CURRENT FILING DATE: 2004-03-04
/ PRIOR APPLICATION NUMBER: 60/164,258
/ PRIOR FILING DATE: 1999-11-09
/ NUMBER OF SEQ ID NOS: 4472
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1780
/ LENGTH: 1155
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic
/ OTHER INFORMATION: amino acid sequence
US-10-793-626-1780
```

```
Query Match          12.1%; Score 78.5; DB 6; Length 1155;
Best Local Similarity 27.3%; Pred. No. 46;
Matches 24; Conservative 12; Mismatches 27; Indels 25; Gaps 2;

QY 53 EKKEEENKPTFDVSKKKN-----PQVNHSQLNESHKEDLQREHSQK 96
Db 100 EVKKVEAPTTSQVSKPANEAVVTNESTKPKTTEAPTVEBSIAETPKTSTTQDSTK 159

QY 97 SDSTKQVTVATVLDKNNISSKSTNNPNK 124
Db 160 NPSLKD-----NLNSSSTTSKSK 178
```

## RESULT 21

```
US-10-485-517-239
/ Sequence 239, Application US/10485517
/ Publication No. US20050256299A1
/ GENERAL INFORMATION:
/ APPLICANT: University of Sheffield
/ APPLICANT: Biosynexus Incorporated
/ APPLICANT: Foster, Simon
/ APPLICANT: Mond, James
/ TITLE OF INVENTION: Antigenic Polypeptides
/ FILE REFERENCE: P100629WO
/ CURRENT APPLICATION NUMBER: US/10/485,517
/ CURRENT FILING DATE: 2004-02-02
/ PRIOR APPLICATION NUMBER: GB 0118825.9
/ PRIOR FILING DATE: 2001-08-02
/ PRIOR APPLICATION NUMBER: GB 0200349.9
/ PRIOR FILING DATE: 2002-01-09
/ NUMBER OF SEQ ID NOS: 424
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 239
/ LENGTH: 405
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-10-485-517-239
```

```
Query Match          11.9%; Score 77.5; DB 6; Length 405;
Best Local Similarity 25.7%; Pred. No. 17;
Matches 22; Conservative 16; Mismatches 33; Indels 3; Gaps 2;

QY 49 KKIEEKEENKPTFDVSKKKNPQVNHSQLN-ESHKEDLQREHSQKSDSTKQVTVATV 107
Db 51 KAAESTKNELNEATISADNQSSDKVDVQQLNQEDNTKNDNQKEMVSSQGNETTSGNKL 110

QY 108 LDKNNISSKSTNN 121
Db 111 IEKESV--QSTTGN 122
```



```
RESULT 22
US-10-873-528-191
; Sequence 191, Application US/10873528
; Publication No. US2005027681A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hanbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 191
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-191

Query Match      11.9%; Score 77.5; DB 6; Length 627;
Best Local Similarity 23.4%; Pred. No. 28;
Matches 30; Conservative 26; Mismatches 59; Indels 13; Gaps 4;

Qy 10 CELEKGY--QFDGWEISGEFGKAGYVINLSKDTFIPKVPKIEEKKKEENKPT----- 62
Db 497 GAMQGWKKANDWFFYKTDGSRAGWIKDKQKWFLEKNGQLLVNGKTPGTYVDSSGA 556

Qy 63 --FDVSKKKON--POVNHQSQNSHR--KEDLQREHSQKSDSTKDVATVLDKNNISK 116
Db 557 WLVDVSIKSTATIKTTSHSEIKESKEVVVKOLENKETSHESVTFNSTSQDLTSSTSS 616

Qy 117 STTNPNK 124
Db 617 ETSVKNSE 624

RESULT 23
US-11-196-475-70
; Sequence 70, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; PRIOR FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR FILING DATE: 1994-04-29
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-70

Query Match      11.8%; Score 77; DB 7; Length 663;
Best Local Similarity 31.3%; Pred. No. 33;
Matches 26; Conservative 20; Mismatches 29; Indels 8; Gaps 4;

Qy 45 KPVFKIEEKKKEENKPTFDVSKKONPQVNHQ-LN-ESHRKEDLQREHSQKSDSTKD 102
Db 378 REIEKQIEIKKND-----ELFKKHDKALDKQLNSKASSKEKIGREEDKELDSKN 432

Qy 103 V-TATVLDKNNISKSTTNPNK 124
Db 433 LEPVSEADKVDKISKNNNEVSK 455

RESULT 24
US-11-196-475-78
; Sequence 78, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; PRIOR FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-78

Query Match      11.8%; Score 77; DB 7; Length 663;
Best Local Similarity 31.3%; Pred. No. 33;
Matches 26; Conservative 20; Mismatches 29; Indels 8; Gaps 4;

Qy 45 KPVFKIEEKKKEENKPTFDVSKKONPQVNHQ-LN-ESHRKEDLQREHSQKSDSTKD 102
Db 378 REIEKQIEIKKND-----ELFKKHDKALDKQLNSKASSKEKIGREEDKELDSKN 432

Qy 103 V-TATVLDKNNISKSTTNPNK 124
Db 433 LEPVSEADKVDKISKNNNEVSK 455

RESULT 25
US-11-188-298-17915
; Sequence 17915, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
```

```
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-70

Query Match      11.8%; Score 77; DB 7; Length 663;
Best Local Similarity 31.3%; Pred. No. 33;
Matches 26; Conservative 20; Mismatches 29; Indels 8; Gaps 4;

Qy 45 KPVFKIEEKKKEENKPTFDVSKKONPQVNHQ-LN-ESHRKEDLQREHSQKSDSTKD 102
Db 378 REIEKQIEIKKND-----ELFKKHDKALDKQLNSKASSKEKIGREEDKELDSKN 432

Qy 103 V-TATVLDKNNISKSTTNPNK 124
Db 433 LEPVSEADKVDKISKNNNEVSK 455

RESULT 24
US-11-196-475-78
; Sequence 78, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; PRIOR FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-78

Query Match      11.8%; Score 77; DB 7; Length 663;
Best Local Similarity 31.3%; Pred. No. 33;
Matches 26; Conservative 20; Mismatches 29; Indels 8; Gaps 4;

Qy 45 KPVFKIEEKKKEENKPTFDVSKKONPQVNHQ-LN-ESHRKEDLQREHSQKSDSTKD 102
Db 378 REIEKQIEIKKND-----ELFKKHDKALDKQLNSKASSKEKIGREEDKELDSKN 432

Qy 103 V-TATVLDKNNISKSTTNPNK 124
Db 433 LEPVSEADKVDKISKNNNEVSK 455

RESULT 25
US-11-188-298-17915
; Sequence 17915, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
```





## RESULT 33

US-11-185-924-16  
; Sequence 16, Application US/11185924  
; Publication No. US20060078945A1  
; GENERAL INFORMATION:  
; APPLICANT: Fisher et al., Larry  
; TITLE OF INVENTION: Complex Formed by Small Integrin-Binding Ligand,  
; TITLE OF INVENTION: N-Linked Glycoproteins (SIBLINGS) and Factor H  
; FILE REFERENCE: 4239-61301-02  
; CURRENT APPLICATION NUMBER: US/11/185,924  
; CURRENT FILING DATE: 2005-07-19  
; PRIOR APPLICATION NUMBER: 09/958,617  
; PRIOR FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: PCT/US00/09349  
; PRIOR FILING DATE: 2000-04-09  
; PRIOR APPLICATION NUMBER: 60/128,468  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 513  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-185-924-16

Query Match 11.5%; Score 75; DB 7; Length 513;  
Best Local Similarity 23.5%; Pred. No. 37;  
Matches 23; Conservative 27; Mismatches 34; Indels 14; Gaps 4;  
  
QY 29 KKDAGVNLKDTIKPVFKIEKKEENKPTFDVSKKK-DNPQVNHSQLNE-----S 82  
Db 385 ESDSHTLHSHKS-----ERERQADSESESLNFSESPEDENSQQGLQSHS 437  
  
QY 83 HRKEDLQREHSQKSDS-TKQVTVTLQKNINSSKSTT 119  
Db 438 SSAESQSESHSEEDSDSQDSRSKEDSNSTESKSS 475

## RESULT 34

US-11-096-568A-32043  
; Sequence 32043, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 32043  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(313)  
; OTHER INFORMATION: Ceres Seq. ID no. 13592023  
US-11-096-568A-32043

Query Match 11.4%; Score 74.5; DB 7; Length 313;  
Best Local Similarity 22.0%; Pred. No. 23;  
Matches 26; Conservative 24; Mismatches 49; Indels 19; Gaps 4;  
  
QY 1 EDFILPVYK-----GELEKGYQPDGWEISGFEKGKDGAGYVINLSKDTF-----IKPVF 48  
Db 196 ELYVFPKLYGVGVELWQVKSGLFDNVLVS-----DPEYAKKLAETWGHKDKAEKAAF 250  
  
QY 49 KKIEKKEENKPTFDVSKKDNQPNVNHSQLNESHKEDLQREHSQKSDSTKQVTVAT 106  
Db 251 DEAEKKREESK--DAPAESDAEEAEDDDNEGDDSDNESKSEBTKAEETKEAET 306

## RESULT 35

US-11-096-568A-32042  
; Sequence 32042, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 32042  
; LENGTH: 425  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(425)  
; OTHER INFORMATION: Ceres Seq. ID no. 13592022  
US-11-096-568A-32042

Query Match 11.4%; Score 74.5; DB 7; Length 425;  
Best Local Similarity 22.0%; Pred. No. 32;  
Matches 26; Conservative 24; Mismatches 49; Indels 19; Gaps 4;  
  
QY 1 EDFILPVYK-----GELEKGYQPDGWEISGFEKGKDGAGYVINLSKDTF-----IKPVF 48  
Db 308 ELYVFPKLYGVGVELWQVKSGLFDNVLVS-----DPEYAKKLAETWGHKDKAEKAAF 362  
  
QY 49 KKIEKKEENKPTFDVSKKDNQPNVNHSQLNESHKEDLQREHSQKSDSTKQVTVAT 106  
Db 363 DEAEKKREESK--DAPAESDAEEAEDDDNEGDDSDNESKSEBTKAEETKEAET 418

## RESULT 36

US-11-096-568A-32041  
; Sequence 32041, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 32041  
; LENGTH: 434  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(434)  
; OTHER INFORMATION: Ceres Seq. ID no. 13592021  
US-11-096-568A-32041

Query Match 11.4%; Score 74.5; DB 7; Length 434;  
Best Local Similarity 22.0%; Pred. No. 33;  
Matches 26; Conservative 24; Mismatches 49; Indels 19; Gaps 4;  
  
QY 1 EDFILPVYK-----GELEKGYQPDGWEISGFEKGKDGAGYVINLSKDTF-----IKPVF 48  
Db 317 ELYVFPKLYGVGVELWQVKSGLFDNVLVS-----DPEYAKKLAETWGHKDKAEKAAF 371  
  
QY 49 KKIEKKEENKPTFDVSKKDNQPNVNHSQLNESHKEDLQREHSQKSDSTKQVTVAT 106  
Db 372 DEAEKKREESK--DAPAESDAEEAEDDDNEGDDSDNESKSEBTKAEETKEAET 427

## RESULT 37

US-11-196-475-68  
; Sequence 68, Application US/11196475  
; Publication No. US20050271682A1  
; GENERAL INFORMATION:  
; APPLICANT: Dattwyler, Raymond J.  
; APPLICANT: Gomes Solecki, Maria J. C.  
; APPLICANT: Luft, Benjamin J.  
; APPLICANT: Dunn, John J.  
; TITLE OF INVENTION: Recombinant Constructs of Borrelia  
; FILE REFERENCE: 2631.1001-011  
; CURRENT APPLICATION NUMBER: US/11/196,475  
; CURRENT FILING DATE: 2005-08-03  
; PRIOR APPLICATION NUMBER: US 08/148,191  
; PRIOR FILING DATE: 1993-11-01  
; PRIOR APPLICATION NUMBER: US 08/235,836  
; PRIOR FILING DATE: 1994-04-29  
; PRIOR APPLICATION NUMBER: US 09/666,017  
; PRIOR FILING DATE: 2000-09-19  
; PRIOR APPLICATION NUMBER: US 60/226,484  
; PRIOR FILING DATE: 2000-08-18  
; PRIOR APPLICATION NUMBER: PCT/US01/24736  
; PRIOR FILING DATE: 2001-08-07  
; NUMBER OF SEQ ID NOS: 213  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 68  
; LENGTH: 693  
; TYPE: PRT  
; ORGANISM: Borrelia burgdorferi  
US-11-196-475-68

Query Match 11.4%; Score 74.5; DB 7; Length 693;  
Best Local Similarity 28.7%; Pred. No. 57;  
Matches 25; Conservative 20; Mismatches 31; Indels 11; Gaps 4;  
QY 36 INLSKDTFKPVFKIEEKEENKPTFDVSKKONQVNHSQLNESHKEDLQRE 91  
DB 316 LDIQDTVREKIQEDIDEINKEKLPKPGDVSPKVKQL---QIKES--LEDLQELKE 370  
QY 92 --EHSQKSDTKDVTATVLDKNNISK 116  
DB 371 TSDENQKREIKQIEIKKSDLELLSK 397

RESULT 38  
US-11-087-099-5463  
; Sequence 5463, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 5463  
; LENGTH: 550  
; TYPE: PRT  
; ORGANISM: Pisum sativum  
US-11-087-099-5463

Query Match 11.4%; Score 74; DB 7; Length 550;  
Best Local Similarity 20.3%; Pred. No. 49;  
Matches 31; Conservative 24; Mismatches 42; Indels 56; Gaps 4;  
QY 19 DGWEISGFGP-----KQDAGYVIN-----LSKDTF 43  
DB 397 EGKELGHPAGSAGKPKIBAYDKDKKGGGLITPAKTNTAADSIVIEPKNSAMDEDT 456  
QY 44 IKVPVKIEEKEENKPTFDVSKKONQVNHSQLNESHKEDLQREHSOKSDTKDV 103  
DB 457 ETPVTGKKKKEK-----KKREKKEKVEDVVEPQEVVKKKKKKKKKSTKA 507

QY 104 TATVLDKN-----NISKSTTNNPNK 124  
DB 508 EVQGDNNLNGEKKKKKKRKSABEEDTTDMPSK 540  
RESULT 39  
US-11-087-099-9570  
; Sequence 9570, Application US/11087099  
; Publication No. US20060041961A1  
; GENERAL INFORMATION:  
; APPLICANT: Abad, Mark S. et al.  
; TITLE OF INVENTION: Genes and Uses for Plant Improvement  
; FILE REFERENCE: 38-21(53450)B EP  
; CURRENT APPLICATION NUMBER: US/11/087,099  
; CURRENT FILING DATE: 2005-03-22  
; NUMBER OF SEQ ID NOS: 12464  
; SEQ ID NO 9570  
; LENGTH: 1758  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
US-11-087-099-9570

Query Match 11.4%; Score 74; DB 7; Length 1758;  
Best Local Similarity 26.4%; Pred. No. 1.9e+02;  
Matches 14; Conservative 16; Mismatches 23; Indels 0; Gaps 0;  
QY 49 KKIEKKEEENKPTFDVSKKONQVNHSQLNESHKEDLQREHSOKSDSTK 101  
DB 1017 KKDMDSEENSEYNTAAEQSPQPKSKINNKKRKKQKQKQAEQAEEK 1069

RESULT 40  
US-10-821-234-1536  
; Sequence 1536, Application US/10821234  
; Publication No. US20050255114A1  
; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Andarmani, Susan  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia  
; FILE REFERENCE: 821A  
; CURRENT APPLICATION NUMBER: US/10/821,234  
; CURRENT FILING DATE: 2004-04-07  
; PRIOR APPLICATION NUMBER: US 60/462,047  
; PRIOR FILING DATE: 2003-04-07  
; NUMBER OF SEQ ID NOS: 1704  
; SOFTWARE: PCT\_SEQ\_genes Version 1.0  
; SEQ ID NO 1536  
; LENGTH: 417  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-821-234-1536

Query Match 11.3%; Score 73.5; DB 6; Length 417;  
Best Local Similarity 19.6%; Pred. No. 39;  
Matches 20; Conservative 25; Mismatches 50; Indels 7; Gaps 2;  
QY 11 ELEKGYQPDGWEISGFGKDKAGYVINLSKDTF--IKVPVKIEEKEENKPTFDVSKK 68  
DB 320 QVKSGLTFDNLIT-----NDEAYAEFGNETGWTKAAEQMKQKQDEQRLKEEEDK 374  
QY 69 KDNQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 110  
DB 375 KRKEEAEADKEDDEDKDE 416

Search completed: April 24, 2006, 15:44:46  
Job time : 10.7092 secs

**THIS PAGE BLANK (user)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 62.1387 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773

Perfect score: 651

Sequence: 1 EDFILPVYKGELEKGYQFDG.....ATVLDKNNISSKSTNNPNK 124

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	100.0	773	4	US-10-067-385-8
2	651	100.0	2119	3	US-09-769-744A-28
3	651	100.0	2140	4	US-10-282-122A-73670
4	651	100.0	2140	5	US-10-472-928-1180
5	648	99.5	637	5	US-10-617-320-3169
6	615	94.5	117	3	US-09-765-272-68
7	615	94.5	117	6	US-11-106-649-68
8	106	16.3	188	5	US-10-691-672A-7
9	101.5	15.6	564	6	US-11-097-143-12723
10	101	15.5	169	5	US-10-691-672A-2
11	101	15.5	647	5	US-10-691-672A-3
12	98	15.1	665	3	US-09-820-843A-107
13	96.5	14.8	707	4	US-10-282-122A-52942
14	93	14.3	869	4	US-10-437-963-12282
15	93	14.3	1529	5	US-10-732-923-8762
16	92.5	14.2	2468	4	US-10-755-889-615
17	92.5	14.2	2468	5	US-10-489-740-216
18	92.5	14.2	2519	5	US-10-450-763-46995
19	92	14.1	1005	4	US-10-437-963-187665
20	91	14.0	470	5	US-10-739-930-6262
21	89.5	13.7	511	4	US-10-289-762-509
22	88	13.5	225	4	US-10-032-585-7829
23	88	13.5	815	5	US-10-496-905-23
24	88	13.5	1980	5	US-10-482-834A-144
25	87.5	13.4	903	4	US-10-282-122A-52328
26	87.5	13.4	1702	3	US-09-839-996-5
27	87.5	13.4	1702	4	US-10-080-505-5

28	87.5	13.4	1702	4	US-10-645-655-5	Sequence 5, Appli
29	87.5	13.4	1702	4	US-10-687-046-5	Sequence 5, Appli
30	87	13.4	593	4	US-10-424-599-174814	Sequence 174814, A
31	86.5	13.3	1373	5	US-10-732-923-16976	Sequence 16976, A
32	86	13.2	145	4	US-10-424-599-149572	Sequence 149572, A
33	86	13.2	835	4	US-10-425-115-238086	Sequence 238086, A
34	86	13.2	1075	4	US-10-437-963-187664	Sequence 187664, A
35	86	13.2	1726	5	US-10-942-522-6	Sequence 6, Appli
36	86	13.2	1980	4	US-10-474-778-6	Sequence 6, Appli
37	86	13.2	1980	5	US-10-482-834A-150	Sequence 150, App
38	86	13.2	1980	5	US-10-942-522-8	Sequence 8, Appli
39	85.5	13.1	884	4	US-10-032-585-7212	Sequence 7212, Ap
40	85.5	13.1	982	4	US-10-425-115-293706	Sequence 293706, A
41	85	13.1	316	4	US-10-424-599-180899	Sequence 180899, A
42	85	13.1	465	4	US-10-724-972A-5599	Sequence 5599, Ap
43	85	13.1	728	4	US-10-282-122A-70944	Sequence 70944, A
44	85	13.1	728	4	US-10-724-972A-4389	Sequence 4389, Ap
45	85	13.1	778	4	US-10-724-972A-5663	Sequence 5663, Ap

ALIGNMENTS

RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 651;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 7.6e-51;		
Matches 124;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	EDFILPVYKGELEKGYQFDGWEISGFSGKGDAGYVNLKSDTFIKPVFKLEEKEENK	60	
Db	650	EDFILPVYKGELEKGYQFDGWEISGFSGKGDAGYVNLKSDTFIKPVFKLEEKEENK	709	
Qy	61	PTFDVSKKQNPQVNHSQLNESHKEDLQREHQSQSDSTKDVATVLDKNNISSKSTTN	120	
Db	710	PTFDVSKKQNPQVNHSQLNESHKEDLQREHQSQSDSTKDVATVLDKNNISSKSTTN	769	
Qy	121	NNPK 124		
Db	770	NNPK 773		

RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins

```
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 651; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 2.5e-50;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKKGADAGYVINLSKDTFIKPVFKKIEEKKBEENK 60
    |||
Db 1962 EDFILPVYKGELEKGYQFDGWEISGPEGKKGADAGYVINLSKDTFIKPVFKKIEEKKBEENK 2021

QY 61 PTFDVSKKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 120
    |||
Db 2022 PTFDVSKKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 2081

QY 121 NPNK 124
    |||
Db 2082 NPNK 2085

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match      100.0%; Score 651; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.6e-50;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKKGADAGYVINLSKDTFIKPVFKKIEEKKBEENK 60
    |||
Db 1983 EDFILPVYKGELEKGYQFDGWEISGPEGKKGADAGYVINLSKDTFIKPVFKKIEEKKBEENK 2042

QY 61 PTFDVSKKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 120
    |||
Db 2043 PTFDVSKKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 2102

QY 121 NPNK 124
    |||
Db 2103 NPNK 2106

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/10472923
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
US-10-472-928-1180

Query Match      100.0%; Score 651; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.6e-50;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKKGADAGYVINLSKDTFIKPVFKKIEEKKBEENK 60
    |||
Db 1983 EDFILPVYKGELEKGYQFDGWEISGPEGKKGADAGYVINLSKDTFIKPVFKKIEEKKBEENK 2042

QY 61 PTFDVSKKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 120
    |||
Db 2043 PTFDVSKKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 2102

QY 121 NPNK 124
    |||
Db 2103 NPNK 2106

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```



```
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
; THERAPEUTICS
;
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-10-617-320-3169

Query Match          99.5%; Score 648; DB 5; Length 637;
Best Local Similarity 99.2%; Pred No. 1.1e-50;
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFKPVFKIEEKKKEENK 60
Db 480 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFKPVFKIEEKKKEENK 539

Qy 61 PTFDVSCKKDPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISKSTTN 120
Db 540 PTFDVSCKKDPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISKSTTN 599

Qy 121 NPNK 124
Db 600 NPNK 603

; RESULT 6
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

```
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match          94.5%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.5e-48;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFKPVFKIEEKKKEENKPTFDVSK 67
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFKPVFKIEEKKKEENKPTFDVSK 60

Qy 68 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISKSTTNPNK 124
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISKSTTNPNK 117

; RESULT 7
US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```



Qy 120 NN 121  
Db 125 NN 126

RESULT 11  
US-10-691-672A-3  
; Sequence 3, Application US/10691672A  
; Publication No. US20050112133A1  
; GENERAL INFORMATION:  
; APPLICANT: DRUILHE, PIERRE  
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND  
; FILE REFERENCE: 02356.0085  
; CURRENT APPLICATION NUMBER: US/10/691,672A  
; CURRENT FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 3.3  
; SEQ ID NO 3  
; LENGTH: 647  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; NAME/KEY: SITE  
; LOCATION: (1)..(647)  
; OTHER INFORMATION: GLURP MSP3 fusion protein  
US-10-691-672A-3

Query Match 15.5%; Score 101; DB 5; Length 647;  
Best Local Similarity 25.4%; Pred. No. 1.2;  
Matches 31; Conservative 26; Mismatches 39; Indels 26; Gaps 5;

Qy 20 GWEISGP--EGKQAG-----YVINLSKDTPIKPVFKKIEKKKEENKPTFDVSKKQNP 72  
Db 489 GWEFGGVPKHKKENMLSHLYVSSKKENISKENDDDVLDS-KBEAEETEEELERKE 547  
Qy 73 QVNHSQLN-----ESHKEDLQREHSQKSDSTKQVATVLDKNNISSKSTT 119  
Db 548 EETSEISEDEEEEEKEBENEKKEQEKEQENNDQKKDMEA-----QNLISKQN 602

Qy 120 NN 121  
Db 603 NN 604

RESULT 12  
US-09-820-843A-107  
; Sequence 107, Application US/09820843A  
; Publication No. US20030039963A1  
; GENERAL INFORMATION:  
; APPLICANT: Council of Scientific and Industrial Research  
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN  
; FILE REFERENCE: Q63915  
; CURRENT APPLICATION NUMBER: US/09/820,843A  
; CURRENT FILING DATE: 2001-03-30  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 107  
; LENGTH: 665  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: hypothetical protein  
; NAME/KEY: misc feature  
; OTHER INFORMATION: GI|3845248  
US-09-820-843A-107

Query Match 15.1%; Score 98; DB 3; Length 665;  
Best Local Similarity 27.1%; Pred. No. 2.3;  
Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;

Qy 13 EKGQFDGWEI--SGPEKKDAGYVINLSKDTPIKPVFKKIEKKKE-----EENK 60  
Db 158 EKGQKQ----DISNSAENKKD-----VKBGVKELEKKKKKEKISDDHKVEENK 201  
Qy 61 PTFD-----VSKKONPQVNHSQLNESHKEDLQREHSQKSDSTKQVATVLDKNNISS 115  
Db 202 KSDDHKEENKSDDHKEENKSDDHKEIEHVKVKEHEDEEE-----DKCEKKS 252

Qy 116 KSTNNPNK 124  
Db 253 ENKKNENK 261

RESULT 13  
US-10-282-122A-52942  
; Sequence 52942, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 52942  
; LENGTH: 707  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (6)..(6)  
; OTHER INFORMATION: X=any amino acid  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (18)..(18)  
; OTHER INFORMATION: X=any amino acid

```
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (29)..(29)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (37)..(37)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (43)..(43)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (54)..(54)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (84)..(84)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (86)..(86)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (359)..(359)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (385)..(385)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (388)..(388)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (396)..(396)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (400)..(400)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (402)..(402)
; OTHER INFORMATION: X=any amino acid
;
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (404)..(404)
; OTHER INFORMATION: X=any amino acid
;
US-10-282-122A-52942
```

```
Query Match          14.8%; Score 96.5; DB 4; Length 707;
Best Local Similarity 25.2%; Pred. No. 3.4;
Matches 34; Conservative 25; Mismatches 53; Indels 23; Gaps 4;

Qy 9 KQELKGYQFGWEISGPEGKADAGVINLSKDTPIKPVFKKIEK----- 54
Db 495 KEEIDSNQDIGVV---EDKDTTDEYDSNKEIDIEPENKSKKKAKLFGFIKKDNEV 551
Qy 55 -KEEN-----KPTFVSKKDNPNQVNHSQLSHRKEDLQREHSQKSDSKDVTATVLD 109
Db 552 EQEENLNDISPDITLDKPVENNQVSKBEEIQENLKE-IKOEPSQHIERSVKIEKPI 610
Qy 110 KNNISSKSTTNPNK 124
Db 611 NNNLDEKVSSNNESK 625
```

```
RESULT 14
US-10-437-963-122282
```

```
; Sequence 122282, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122282
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRM4530_25224C.1.pep
US-10-437-963-122282

Query Match          14.3%; Score 93; DB 4; Length 869;
Best Local Similarity 45.1%; Pred. No. 9.2;
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;

Qy 64 DVSKKDNPNQVNHSQLSHRKEDLQREHSQKSDSKDVTATVLDKNN 112
Db 617 DASKRKDNHQSNNL--SHRDEDPTRKKKQKTNATSDACAQBVVTEKNN 665
```

```
RESULT 15
US-10-732-923-8762
; Sequence 8762, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8762
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1529)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-8762
```

```
Query Match          14.3%; Score 93; DB 5; Length 1529;
Best Local Similarity 27.2%; Pred. No. 18;
Matches 28; Conservative 19; Mismatches 30; Indels 26; Gaps 5;

Qy 37 NLSKDT-----PIKPVFKKIEKKKEENKPTFDVSKKDNPNQVNHSQL-----NES 82
Db 356 NIISDTQCIKPIKVIINSEYKNEKKNEKK-----NEKINTIHYSESISKNSDNEQ 409
Qy 83 H-----RKEDLQREHSQKSDSKDVTATVLDKNN--ISSKSTT 119
Db 410 HPFLSKLRNVKKEKKKKKKTKIKTVIAQKNKTVIAQKNKT 452
```

```
RESULT 16
US-10-755-889-615
; Sequence 615, Application US/10755889
```

; Publication No. US20040171823A1  
; GENERAL INFORMATION:  
; APPLICANT: Bristol-Myers Squibb Company  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB  
; FILE REFERENCE: D0284 NP  
; CURRENT APPLICATION NUMBER: US/10/755,889  
; CURRENT FILING DATE: 2004-01-13  
; PRIOR APPLICATION NUMBER: U.S. 60/440,068  
; PRIOR FILING DATE: 2003-01-14  
; PRIOR APPLICATION NUMBER: U.S. 60/469,757  
; PRIOR FILING DATE: 2003-05-12  
; NUMBER OF SEQ ID NOS: 823  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 615  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-755-889-615

Query Match 14.2%; Score 92.5; DB 4; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 36;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
  
QY 38 LSKDTFIKVPFKIEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE-----E 92  
Db 638 VKKETKVKP-----EDKKEEKPKKKEVAKKEDKTPI---KKEKPKKEEVKKEVKKKEIK 689  
  
QY 93 HSQKSDSTKDV 103  
Db 690 KEEKKEPKKEV 700

RESULT 17  
US-10-489-740-216  
; Sequence 216, Application US/10489740  
; Publication No. US20050112574A1  
; GENERAL INFORMATION:  
; APPLICANT: Biomedics Limited  
; TITLE OF INVENTION: P9  
; FILE REFERENCE: Angiogenesis PCT  
; CURRENT APPLICATION NUMBER: US/10/489,740  
; CURRENT FILING DATE: 2004-03-15  
; NUMBER OF SEQ ID NOS: 216  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 216  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-489-740-216

Query Match 14.2%; Score 92.5; DB 5; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 36;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
  
QY 38 LSKDTFIKVPFKIEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE-----E 92  
Db 638 VKKETKVKP-----EDKKEEKPKKKEVAKKEDKTPI---KKEKPKKEEVKKEVKKKEIK 689  
  
QY 93 HSQKSDSTKDV 103  
Db 690 KEEKKEPKKEV 700

RESULT 18  
US-10-450-763-46995  
; Sequence 46995, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
; FILE REFERENCE: 790C1F3/US  
; CURRENT APPLICATION NUMBER: US/10/450,763

; CURRENT FILING DATE: 2003-06-11  
; PRIOR APPLICATION NUMBER: PCT/US01/08631  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: 09/649,167  
; PRIOR FILING DATE: 2000-08-23  
; NUMBER OF SEQ ID NOS: 60736  
; SOFTWARE: Custom  
; SEQ ID NO 46995  
; LENGTH: 2519  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (1040)..(1091)  
; OTHER INFORMATION: Neuromodulin (GAP-43) proteins domain identified by eMATRIX,  
; OTHER INFORMATION: accession number BL00412D, p-value=2.432e-11, raw score of 16.54  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (1919)..(2122)  
; OTHER INFORMATION: Neuraxin and MAP1B proteins domain identified by PFam,  
; OTHER INFORMATION: accession name MAP1B\_neuraxin, E-value=1.9e-59, PFam score of 15.54  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(2519)  
; OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
US-10-450-763-46995

Query Match 14.2%; Score 92.5; DB 5; Length 2519;  
Best Local Similarity 31.0%; Pred. No. 37;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
  
QY 38 LSKDTFIKVPFKIEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE-----E 92  
Db 689 VKKETKVKP-----EDKKEEKPKKKEVAKKEDKTPI---KKEKPKKEEVKKEVKKKEIK 740  
  
QY 93 HSQKSDSTKDV 103  
Db 741 KEEKKEPKKEV 751

RESULT 19  
US-10-437-963-187665  
; Sequence 187665, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 187665  
; LENGTH: 1005  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_84346C.1.pap  
US-10-437-963-187665

Query Match 14.1%; Score 92; DB 4; Length 1005;  
Best Local Similarity 26.4%; Pred. No. 14;  
Matches 32; Conservative 22; Mismatches 41; Indels 26; Gaps 4;

```

Qy 25 GFGEKQDAGVYINLSKDTFIKPVFKIIEKKEEENKPTFDVSKK--KDPQV--NHSQLINE 81
Db 35 GSEKEMSGKNIKSIKETGTTGQSKELQKGSKSRKSTKDKSKKKQKMDTQVPTNAEEFHK 94
Qy 82 SH-----RKSE-----DLQREHSQKSDSTKQVTAIVLDKKNISSKST 118
Db 95 EYTKVIRKESRTSSSIEQVIGTSSIOEMETNEQVKSQKTSKOMTQVPANAEGIRKEYT 154
Qy 119 T 119
Db 155 T 155

RESULT 20
US-10-739-930-6262
; Sequence 6262, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: KOVALIC, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 6262
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C271270_1.p
US-10-739-930-6262

Query Match 14.0%; Score 91; DB 5; Length 470;
Best Local Similarity 19.3%; Pred. No. 6.7;
Matches 29; Conservative 29; Mismatches 58; Indels 34; Gaps 4;

Qy 8 YKGELEKGYQPDGWEISGFE-----GKIDAGVYINLSKDTFIKPVFKIIEKKEEEN 59
Db 100 YVODLARRIRYDE-EATGQSAQRIDHPNQKNVGITEKAFENSPIEETSHRVDDNKRINN 158
Qy 60 KPTFDVSKKKN-----PQVNHSQLNE-----SHRKEDLQREHS 94
Db 159 QKNFTAAKSSNAVSRVSFGADHKRAEYGVGKPMENDQVQTESAKSHRKENVTYKSEK 218
Qy 95 QKSDSTKDVTAIVLDKKNISSKSTTNPNK 124
Db 219 RDQGVKTEAKDKORNKKEKTESINK 248

RESULT 21
US-10-289-762-509
; Sequence 509, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Grifflab, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fr
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; CURRENT FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 509
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-10-289-762-509

Query Match 13.7%; Score 89.5; DB 4; Length 511;
Best Local Similarity 24.5%; Pred. No. 10;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

```



```
;
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-839-996-5

Query Match 13.4%; Score 87.5; DB 3; Length 1702;
Best Local Similarity 27.2%; Pred. No. 65;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 36 INLSKDTFKVPFKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS- 94
Db 1296 INTGSATAITETAESKDPQTETAASTEDASQHKANTVADNSVANNSESESPKRRRSI 1355

Qy 95 --QKSDSTKDVATVLDKNNISSKSTNNPNK 124
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 27
US-10-080-505-5
; Sequence 5, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1702
; TYPE: PRP
; ORGANISM: Haemophilus influenzae
US-10-080-505-5

Query Match 13.4%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 65;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 36 INLSKDTFKVPFKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS- 94
Db 1296 INTGSATAITETAESKDPQTETAASTEDASQHKANTVADNSVANNSESESPKRRRSI 1355

Qy 95 --QKSDSTKDVATVLDKNNISSKSTNNPNK 124
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 28
US-10-645-655-5
; Sequence 5, Application US/10645655
; Publication No. US20040063908A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; PROTEIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/645,655
; FILING DATE: 20-Aug-2003
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
;
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NC: 5:
US-10-645-655-5

Query Match 13.4%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 65;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 36 INLSKDTFKVPFKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS- 94
Db 1296 INTGSATAITETAESKDPQTETAASTEDASQHKANTVADNSVANNSESESPKRRRSI 1355

Qy 95 --QKSDSTKDVATVLDKNNISSKSTNNPNK 124
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 29
US-10-687-046-5
; Sequence 5, Application US/10687046
; Publication No. US20040157241A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
```







QY 17 QPDGWEISGPEGKDGAGYVINLSKDTFIKVPFKIEEKEENKPTFDVSKKDNQVNH 76  
Db 993 EMNNLQISVIRIKKGVAWT-CLKVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1048  
QY 77 SOLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPN 123  
Db 1049 TGA-DIHRNGDFQKNGGTTSGIGSSVEKYIIDEDHM---SPINNPN 1091

## RESULT 37

US-10-482-834A-150  
; Sequence 150, Application US/10482834A  
; Publication No. US20050074764A1

; GENERAL INFORMATION:  
; APPLICANT: Mulley, John Charles  
; APPLICANT: Harkin, Louise Anne  
; APPLICANT: Dibbens, Michelle  
; APPLICANT: Wallace, Robyn  
; APPLICANT: Phillips, Hillary Ann  
; APPLICANT: Heron, Sara Elizabeth  
; APPLICANT: Berkovic, Samuel Frank  
; APPLICANT: Scheffer, Ingrid Eileen  
; APPLICANT: Bionomics Limited  
; TITLE OF INVENTION: MUTATIONS IN ION CHANNELS  
; FILE REFERENCE: 1386/17  
; CURRENT APPLICATION NUMBER: US/10/482,834A  
; CURRENT FILING DATE: 2004-01-02  
; NUMBER OF SEQ ID NOS: 173  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 150  
; LENGTH: 1980  
; TYPE: PRT  
; ORGANISM: Homo sapiens

## US-10-482-834A-150

Query Match 13.2%; Score 86; DB 5; Length 1980;  
Best Local Similarity 27.1%; Pred. No. 1.1e+02;  
Matches 29; Conservative 21; Mismatches 49; Indels 8; Gaps 4;  
QY 17 QPDGWEISGPEGKDGAGYVINLSKDTFIKVPFKIEEKEENKPTFDVSKKDNQVNH 76  
Db 993 EMNNLQISVIRIKKGVAWT-CLKVHAFMQAHFKQCE---ADEVKPLDELYEKKANCIAH 1048

QY 77 SOLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPN 123

Db 1049 TGA-DIHRNGDFQKNGGTTSGIGSSVEKYIIDEDHM---SPINNPN 1091

## RESULT 38

US-10-942-522-8  
; Sequence 8, Application US/10942522  
; Publication No. US20050112633A1

; GENERAL INFORMATION:  
; APPLICANT: Amour, Christopher D.  
; APPLICANT: Castle, John C.  
; APPLICANT: Garrett-Engle, Philip W.  
; APPLICANT: Kan, Zhengyan  
; APPLICANT: Loerch, Patrick M.  
; APPLICANT: Tsinoemas, Nicholas F.  
; TITLE OF INVENTION: ALTERNATIVELY SPLICED ISOFORMS OF SODIUM  
; TITLE OF INVENTION: CHANNEL, VOLTAGE GATED, TYPE VIII, ALPHA (SCN8A)  
; FILE REFERENCE: RS0214  
; CURRENT APPLICATION NUMBER: US/10/942,522  
; CURRENT FILING DATE: 2004-09-16  
; PRIOR FILING DATE: 2004-09-15  
; PRIOR FILING DATE: 2003-09-17  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 1980  
; TYPE: PRT  
; ORGANISM: Homo sapiens

## US-10-942-522-8

OTHER INFORMATION: Clone ID: MRT4577\_30943C.1.pap

Query Match 13.2%; Score 86; DB 5; Length 1980;  
Best Local Similarity 26.2%; Pred. No. 1.1e+02;  
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;

QY 17 QPDGWEISGPEGKDGAGYVINLSKDTFIKVPFKIEEKEENKPTFDVSKKDNQVNH 76  
Db 993 EMNNLQISVIRIKKGVAWT-CLKVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1048

QY 77 SOLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPN 123

Db 1049 TGA-DIHRNGDFQKNGGTTSGIGSSVEKYIIDEDHM---SPINNPN 1091

## RESULT 39

US-10-032-585-7212  
; Sequence 7212, Application US/10032585  
; Publication No. US20030180953A1

; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 7212  
; LENGTH: 884  
; TYPE: PRT  
; ORGANISM: Candida albicans

## US-10-032-585-7212

Query Match 13.1%; Score 85.5; DB 4; Length 884;  
Best Local Similarity 25.0%; Pred. No. 46;  
Matches 26; Conservative 23; Mismatches 46; Indels 9; Gaps 3;

QY 24 SGFEGKDGAGYVINLSKDTFIKVPFKIE---EKKEENKPTFDVSK--KKDNQVNH 78

Db 14 SGQEVKNTNTSLSKEEKLRRRQLELWRQKQQQQQEQEERVQNKAKKTEDSTNNTS 73

QY 79 LNESHRKEDLQRE---HSQKSDSTKDVATVLDKNNISSKST 118

Db 74 TEEVKKSRQRIEEMKGRARLQKQATKEKTTITIKKVSQTT 117

## RESULT 40

US-10-425-115-293706  
; Sequence 293706, Application US/10425115  
; Publication No. US20040214272A1

; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 293706  
; LENGTH: 982  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(982)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_30943C.1.pap

US-10-425-115-293706

```
Query Match      13.1%; Score 85.5; DB 4; Length 982;
Best Local Similarity 26.7%; Pred. No. 52;
Matches 35; Conservative 18; Mismatches 49; Indels 29; Gaps 5;

QY      20 GWEISGFEGKKDAGYVINLSKDTFIKPVFKK-----IEEKKEEENKPTPDVSKKKON-PQV 74
Db      110 GSRLSKFEKKCDEGFLLEYSSNSKAYRVFNKTHGIIIEAYDVEFPDKTNGSQDESDNLDDV 169

QY      75 NHSQL-----NESHKED-----LOREEH-SQKSDSTKDVATVLDK 110
Db      170 GGTQLINAMKTAIGEIKPKEDDENSVVIPSSSTLNEDHQSOQHDETMTHDQGTSR 229

QY      111 NNISKSSTNN 121
Db      230 HSVPPNASTSN 240
```

Search completed: April 24, 2006, 15:43:31  
Job time : 64.1387 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 19.2796 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773  
Perfect score: 651  
Sequence: 1 EDFILPVYKGELEKGYQFDG.....ATVLDKNNISSKSTNNPNK 124

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5-COMB.pdp.\*  
2: /cgn2\_6/ptodata/1/iaa/6-COMB.pdp.\*  
3: /cgn2\_6/ptodata/1/iaa/H-COMB.pdp.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pdp.\*  
5: /cgn2\_6/ptodata/1/iaa/RE-COMB.pdp.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	100.0	773	2	US-09-590-991-8
2	648	99.5	637	2	US-09-107-433-3169
3	648	99.5	2138	2	US-09-583-110-5274
4	615	94.5	117	2	US-08-961-083-68
5	615	94.5	117	2	US-09-536-784-68
6	615	94.5	117	2	US-09-765-271-68
7	615	94.5	117	2	US-09-765-272A-68
8	96	14.7	347	2	US-09-248-796A-16224
9	95	14.6	348	2	US-09-538-092-1316
10	92.5	14.2	2468	2	US-09-976-594-726
11	92.5	14.2	2468	2	US-09-538-092-1135
12	92.5	14.2	2522	2	US-09-949-016-10237
13	89.5	13.7	511	2	US-09-198-452A-509
14	89.5	13.7	511	2	US-09-438-185A-475
15	87.5	13.4	1702	2	US-08-296-791-5
16	87.5	13.4	1702	2	US-09-839-996-5
17	87.5	13.4	1702	2	US-10-080-505-5
18	87.5	13.4	1702	2	US-10-645-655-5
19	87.5	13.4	1702	4	PCT-US95-10661A-5
20	86	13.2	243	2	US-09-248-796A-20306
21	86	13.2	1989	2	US-09-949-016-10076
22	85.5	13.1	109	2	US-09-248-796A-24668
23	85	13.1	465	2	US-09-134-001C-3856
24	85	13.1	472	2	US-09-710-279-658
25	85	13.1	720	2	US-09-710-279-2058
26	85	13.1	728	2	US-09-134-001C-4968
27	85	13.1	746	2	US-09-710-279-652

28	85	13.1	778	2	US-09-134-001C-3868	Sequence 3868, Ap
29	85	13.1	1976	2	US-09-024-020B-9	Sequence 9, Appli
30	85	13.1	1976	2	US-09-425-043-9	Sequence 9, Appli
31	85	13.1	1978	2	US-09-024-020B-3	Sequence 3, Appli
32	85	13.1	1978	2	US-09-425-043-3	Sequence 3, Appli
33	85	13.1	1988	2	US-09-024-020B-4	Sequence 4, Appli
34	85	13.1	1988	2	US-09-425-043-4	Sequence 4, Appli
35	84.5	13.0	654	2	US-10-172-502-10	Sequence 10, Appli
36	84	12.9	817	2	US-09-248-796A-20276	Sequence 20276, A
37	83.5	12.8	461	2	US-09-949-016-8508	Sequence 8508, Ap
38	83.5	12.8	476	2	US-09-248-796A-15008	Sequence 15008, A
39	83	12.7	278	2	US-09-949-016-10508	Sequence 10508, A
40	82.5	12.7	243	2	US-09-248-796A-21334	Sequence 21334, A
41	82	12.6	1024	2	US-09-270-767-44973	Sequence 44973, A
42	81	12.4	582	2	US-09-147-405B-13	Sequence 13, Appl
43	81	12.4	593	2	US-09-147-405B-11	Sequence 11, Appl
44	81	12.4	1092	2	US-09-147-405B-15	Sequence 15, Appl
45	80	12.3	232	2	US-09-248-796A-23825	Sequence 23825, A

ALIGNMENTS

RESULT 1

US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%	Score 651;	DB 2;	Length 773;
Best Local Similarity	100.0%;	Pred. No. 1.3e-62;		
Matches 124;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	EDFILPVYKGELEKGYQFDGWEISGFGKQDAGVYINLSKDTPIKVPVKLEEKEENK	60	
Db	650	EDFILPVYKGELEKGYQFDGWEISGFGKQDAGVYINLSKDTPIKVPVKLEEKEENK	709	
Qy	61	PTFDVSKKQNPQVNHSQLNESHKREDIQRHEHSQKSDSTKDVATVLDKNNISSKSTTN	120	
Db	710	PTFDVSKKQNPQVNHSQLNESHKREDIQRHEHSQKSDSTKDVATVLDKNNISSKSTTN	769	
Qy	121	NNPK 124		
Db	770	NNPK 773		

RESULT 2

US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; FOR DIAG



```
Db      1 YKGELEKGYQFGWEISGEGKDGAGYVNLKDTFKFPVKIIEKKKEENKPTFDVSK 60
Qy      68 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 124
Db      61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 5
US-09-536-784-68
; Sequence 68, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match      94.5%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 9.1e-60;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      8 YKGELEKGYQFGWEISGEGKDGAGYVNLKDTFKFPVKIIEKKKEENKPTFDVSK 67
Db      1 YKGELEKGYQFGWEISGEGKDGAGYVNLKDTFKFPVKIIEKKKEENKPTFDVSK 60

Qy      68 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 124
Db      61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 6
US-09-765-271-68
; Sequence 68, Application US/09765271
; Patent No. 6887663
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452

Db      1 YKGELEKGYQFGWEISGEGKDGAGYVNLKDTFKFPVKIIEKKKEENKPTFDVSK 60
Qy      68 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 124
Db      61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 7
US-09-765-272A-68
; Sequence 68, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
```

/ APPLICATION NUMBER: US/09/765,272A  
/ FILING DATE: 22-Jan-2001  
/ CLASSIFICATION: <Unknown>  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: 08/961,083  
/ FILING DATE: OCT-30-1997  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Lin J. Hymel  
/ REGISTRATION NUMBER: 45,414  
/ REFERENCE/DOCKET NUMBER: PB340P2C2  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (301) 610-5790  
/ TELEFAX: (301) 309-8439  
/ INFORMATION FOR SEQ ID NO: 68:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 117 amino acids  
/ TYPE: amino acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: protein  
/ SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-765-272A-68  
  
Query Match 94.5%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 9.1e-60;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 8 YGELEKGYQFGWISGEGKDGAGYVNLISKDTFIKPVFKKIEEKEENKPTFDVSK 67  
Db 1 YGELEKGYQFGWISGEGKDGAGYVNLISKDTFIKPVFKKIEEKEENKPTFDVSK 60  
  
QY 68 KDNQPNVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 124  
Db 61 KDNQPNVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 117  
  
RESULT 8  
US-09-248-796A-16224  
/ Sequence 16224, Application US/09248796A  
/ Patent No. 6747137  
/ GENERAL INFORMATION:  
/ APPLICANT: Keith Weinstock et al  
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
/ FILE REFERENCE: 107196.132  
/ CURRENT APPLICATION NUMBER: US/09/248,796A  
/ CURRENT FILING DATE: 1999-02-12  
/ PRIOR APPLICATION NUMBER: US 60/074,725  
/ PRIOR FILING DATE: 1998-02-13  
/ PRIOR APPLICATION NUMBER: US 60/096,409  
/ PRIOR FILING DATE: 1998-08-13  
/ NUMBER OF SEQ ID NOS: 28208  
/ SEQ ID NO 16224  
/ LENGTH: 347  
/ TYPE: PRT  
/ ORGANISM: Candida albicans  
US-09-248-796A-16224  
  
Query Match 14.7%; Score 96; DB 2; Length 347;  
Best Local Similarity 23.3%; Pred. No. 0.026;  
Matches 28; Conservative 22; Mismatches 46; Indels 24; Gaps 3;  
  
QY 11 ELEKGYQFGWISGEGKDGAGYVNLISKDTFIKPVFKKIEEKEENKPTFDVSKKD 70  
Db 238 DLESESEIGWE-----DDLEENTRTGPFVKTLQDLREB-----WKAKEKQA 279  
  
QY 71 NQPNVNSQLN-----ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 124  
Db 280 NPKKEENLNQRPVAKQKQPNSTKKQTKQKQTKKITKPKTSKRMLEGISTSNINK 339  
  
RESULT 9  
US-09-538-092-1316

/ Sequence 1316, Application US/09538092  
/ Patent No. 6753314  
/ GENERAL INFORMATION:  
/ APPLICANT: Giot, Loic  
/ APPLICANT: Mansfield, Traci A.  
/ TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
/ FILE REFERENCE: 15966-542  
/ CURRENT APPLICATION NUMBER: US/09/538,092  
/ CURRENT FILING DATE: 2000-03-29  
/ PRIOR APPLICATION NUMBER: 60/127,352  
/ PRIOR FILING DATE: 1999-04-01  
/ PRIOR APPLICATION NUMBER: 60/178,965  
/ PRIOR FILING DATE: 2000-02-01  
/ NUMBER OF SEQ ID NOS: 1387  
/ SOFTWARE: CurapatSeqFormatter Version 0.9  
/ SEQ ID NO 1316  
/ LENGTH: 348  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (0)...(0)  
/ OTHER INFORMATION: Polypeptide Accession Number Q14093  
US-09-538-092-1316  
  
Query Match 14.6%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 0.033;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;  
  
QY 9 KGELEKGYQFGWISGEGKDGAGYVNLISKDTFIKPVFKKIEEKEENKPTFDV 65  
Db 205 ESEGEKG-----GTEKDSKKGKDS-----KKGKDSAIELQAVKADKKDCKDANKGDE 256  
  
QY 66 SK--KDNQPNVNSQLN-----ESHRKEDLQREHSQKSDSTKD---VTATVLDKNNI 113  
Db 257 SKDAKDAKEIKGKDKKPKPSSTDSKDVKKK---SKDATKDAKVAKDKTEKESA 313  
  
QY 114 SSK 116  
Db 314 DSK 316  
  
RESULT 10  
US-09-976-594-726  
/ Sequence 726, Application US/09976594  
/ Patent No. 6673549  
/ GENERAL INFORMATION:  
/ APPLICANT: Furness, Michael  
/ APPLICANT: Buchbinder, Jenny  
/ TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
/ FILE REFERENCE: PA-0041 US  
/ CURRENT APPLICATION NUMBER: US/09/976,594  
/ CURRENT FILING DATE: 2001-10-12  
/ PRIOR APPLICATION NUMBER: 60/240,409  
/ PRIOR FILING DATE: 2000-10-12  
/ NUMBER OF SEQ ID NOS: 1143  
/ SOFTWARE: PERL Program  
/ SEQ ID NO 726  
/ LENGTH: 2468  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ OTHER INFORMATION: Incyte ID No. 6673549 4295277CDI  
US-09-976-594-726  
  
Query Match 14.2%; Score 92.5; DB 2; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 0.83;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
  
QY 38 LSKDTFIKPVFKKIEEKEENKPTFDVSKKDNQPNVNSQLNESHKEDLQRE-----E 92  
Db 638 VKKETVKVP-----EDKKEEKEKPKKEVAKKEDKTPI---KKKEKPKKEEVKKEVKEIK 689









```
Query Match          13.2%; Score 86; DB 2; Length 243;
Best Local Similarity 28.0%; Pred. No. 0.2;
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;

Qy 16 YQFGWEISGFGKDGAGVINLSKDTTPKVPFK-----IEKKKEENKPTFDVSKKK-- 69
Db 98 YDDDDPEFGESSGAAKELNLSQAIKEWKQRDLIEIEREKLSKKKEIEIEAKS 157

Qy 70 --DNPQVNHSQLNESHKEDLOREH--SQKSDSTKDVTTATVLDKNN 112
Db 158 TIDDPYENYKRDNHQKLEQEKIFSKRDDFLK--RGLWLDVRN 202

RESULT 21
US-09-949-016-10076
; Sequence 10076, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: PsetSeq for Windows Version 4.0
; SEQ ID NO 10076
; LENGTH: 1989
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10076

Query Match          13.2%; Score 86; DB 2; Length 1989;
Best Local Similarity 26.2%; Pred. No. 3.2;
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;

Qy 17 QPFGWEISGFGKDGAGVINLSKDTTPKVPFKIEKKKEENKPTFDVSKKKQNPQNH 76
Db 1002 EWNLIQISVIRIKKGVAT-CLKVAFMQAHFK---QREAEVPLDLYEKKANCIAH 1057

Qy 77 SOLNESHKEDLOREHSQKSDSTKDVTTATVLDKNNISKSTNNPN 123
Db 1058 TGA-DIHRNGDFQKNGTGTSGIGSSVEKYIIDEDHM---SPINNPN 1100

RESULT 22
US-09-248-796A-24668
; Sequence 24668, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; WITH HUMAN DISEASE, METHODS OF DETECTION AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 24668
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-24668
```

```
Query Match          13.1%; Score 85.5; DB 2; Length 109;
Best Local Similarity 32.9%; Pred. No. 0.079;
Matches 27; Conservative 14; Mismatches 26; Indels 15; Gaps 4;

Qy 52 BEKKKEEN-----KPTFDVSK--IKDNPOVNHSQLNESHKEDLOREHS-QKSDSTK 101
Db 11 DDDEEENTKDKNGKEDIDRNKSKEDNSNANSTQAVKNKLETKLKENEHSDEKSDPTK 70

Qy 102 DVTATVLDKNNISKSTNNPN 123
Db 71 ENS-----KDGKVSKEKNTTAN 87

RESULT 23
US-09-134-001C-3856
; Sequence 3856, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3856
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3856

Query Match          13.1%; Score 85; DB 2; Length 465;
Best Local Similarity 32.6%; Pred. No. 0.61;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

Qy 51 IEKKKEENKPTFDVSKKKQNPQVNHSQLNESHKEDLO-----REEHSQKSDS--- 99
Db 3 MEENKQPNKE--NMSNKDDNA-----THLNDSHREDELEFRNKNARQRRRRIDNQSK 56

Qy 100 TKDVTAT-----VLDKNNISKSTNNPNK 124
Db 57 EKDATSTQSQLETKPMDKFDINHKS---HNQNK 86

RESULT 24
US-09-710-279-658
; Sequence 658, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-658

Query Match          13.1%; Score 85; DB 2; Length 472;
```

Best Local Similarity 32.6%; Pred. No. 0.62;  
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;  
QY 51 IEKKEEENKPTFDVSKKKD----NPQVNHSQLNESHKEDLQREHSHKSDSTK----D 102  
Db 46 LEEEQIKALDKKFKASQAKDTNKQNTQNNHQSNNKQNSNDKEKQSQNNKSKPTKKKQON 105  
QY 3 MEENKQPNKB--NMSNKDDNA----TLANDSHRNEDLELFRNNKNAQRERRRIDNQSK 56  
Db 100 TKQVAT-----VLDKNNISSKSTTNNPK 124  
QY 57 EKDATSTOSQLETKPMDKFLDNHKS--HNQNK 86  
Db

RESULT 25  
US-09-710-279-2058  
; Sequence 2058, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2058  
; LENGTH: 720  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-2058

Query Match 13.1%; Score 85; DB 2; Length 720;  
Best Local Similarity 28.0%; Pred. No. 1.1;  
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;  
QY 51 IEKKEEENKPTFDVSKKKD----NPQVNHSQLNESHKEDLQREHSHKSDSTK----D 102  
Db 38 LEEEQIKALDKKFKASQAKDTNKQNTQNNHQSNNKQNSNDKEKQSQNNKSKPTKKKQON 97  
QY 103 VTATVLDKNNISSKSTTNNPK 124  
Db 98 NKGQONKNNKTNKQNNK 119

RESULT 26  
US-09-134-001C-4968  
; Sequence 4968, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4968  
; LENGTH: 728  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4968

Query Match 13.1%; Score 85; DB 2; Length 728;  
Best Local Similarity 28.0%; Pred. No. 1.1;  
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 51 IEKKEEENKPTFDVSKKKD----NPQVNHSQLNESHKEDLQREHSHKSDSTK----D 102  
Db 46 LEEEQIKALDKKFKASQAKDTNKQNTQNNHQSNNKQNSNDKEKQSQNNKSKPTKKKQON 105  
QY 103 VTATVLDKNNISSKSTTNNPK 124  
Db 106 NKGQONKNNKTNKQNNK 127

RESULT 27  
US-09-710-279-652  
; Sequence 652, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 652  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-652

Query Match 13.1%; Score 85; DB 2; Length 746;  
Best Local Similarity 26.4%; Pred. No. 1.1;  
Matches 37; Conservative 19; Mismatches 42; Indels 42; Gaps 8;  
QY 4 ILPVYK-----GELEKGYQFDGW---EISQFE-----GKKDAGYVIN--LSKDTPIK 45  
Db 615 VLPHSKVMLMTDGBLTWP-DMTGWTKEVDLTKLVSTKGVFTVNSISKQIIK 673  
QY 46 PVFKIEEKEENKPTFDVS-----KKQNPQVNHSQLNESHKEDLQREHSHKSDSTK 101  
Db 674 -----NKKIEVLSABDTDDQEKTDDESDSNKSKDKADEHDSNTSSSTK 720  
QY 102 DVTATVLDKNNISSKSTTNN 121  
Db 721 N-----DKSNADSKNDSDD 734

RESULT 28  
US-09-134-001C-3868  
; Sequence 3868, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3868  
; LENGTH: 778  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3868

Query Match 13.1%; Score 85; DB 2; Length 778;



APPLICANT: DIETRICH, PAUL S.  
APPLICANT: FISH, LINDA M.  
APPLICANT: HERMAN, RONALD C.  
APPLICANT: SANGAMESWARAN, LAKSHMI  
TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JANET PAULINE CLARK  
STREET: 3401 HILLVIEW AVENUE, MS A2-250  
CITY: PALO ALTO  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 94304-1397  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/024,020B  
FILING DATE: 16-FEB-1998  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/039,447  
FILING DATE: 26-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: CLARK, JANET P.  
REGISTRATION NUMBER: 34,799  
REFERENCE/DOCKET NUMBER: R0020B-REG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 852-3097  
TELEFAX: (650) 855-5322  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1978 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-024-020B-3  
Query Match 13.1%; Score 85; DB 2; Length 1978;  
Best Local Similarity 25.2%; Pred. No. 4.1;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;  
QY 17 QFDGWEISGPEKKDAGYVNLKDTPIKPVFKKIEKKKEENKPTFDVSKKONPQVNH 76  
Db 991 EMNNLQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1046  
QY 77 SOLNESHKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNNPN 123  
Db 1047 TGV-DIHRNGDFQKNGGTTSGIGSSVEKYIIDEDHM---SFINNPN 1089  
RESULT 32  
US-09-425-043-3  
Sequence 3, Application US/09425043  
Patent No. 6335172  
GENERAL INFORMATION:  
APPLICANT: DELGADO, STEPHEN G.  
APPLICANT: DIETRICH, PAUL S.  
APPLICANT: FISH, LINDA M.  
APPLICANT: HERMAN, RONALD C.  
APPLICANT: SANGAMESWARAN, LAKSHMI  
TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JANET PAULINE CLARK  
STREET: 3401 HILLVIEW AVENUE, MS A2-250  
CITY: PALO ALTO  
STATE: CA

COUNTRY: U.S.A.  
ZIP: 94304-1397  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/425,043  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/024,020  
FILING DATE: 16-FEB-1998  
APPLICATION NUMBER: US 60/039,447  
FILING DATE: 26-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: CLARK, JANET P.  
REGISTRATION NUMBER: 34,799  
REFERENCE/DOCKET NUMBER: R0020B-REG  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 852-3097  
TELEFAX: (650) 855-5322  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1978 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-425-043-3  
Query Match 13.1%; Score 85; DB 2; Length 1978;  
Best Local Similarity 25.2%; Pred. No. 4.1;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;  
QY 17 QFDGWEISGPEKKDAGYVNLKDTPIKPVFKKIEKKKEENKPTFDVSKKONPQVNH 76  
Db 991 EMNNLQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1046  
QY 77 SOLNESHKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNNPN 123  
Db 1047 TGV-DIHRNGDFQKNGGTTSGIGSSVEKYIIDEDHM---SFINNPN 1089  
RESULT 33  
US-09-024-020B-4  
Sequence 4, Application US/09024020B  
Patent No. 6030810  
GENERAL INFORMATION:  
APPLICANT: DELGADO, STEPHEN G.  
APPLICANT: DIETRICH, PAUL S.  
APPLICANT: FISH, LINDA M.  
APPLICANT: HERMAN, RONALD C.  
APPLICANT: SANGAMESWARAN, LAKSHMI  
TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JANET PAULINE CLARK  
STREET: 3401 HILLVIEW AVENUE, MS A2-250  
CITY: PALO ALTO  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 94304-1397  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/024,020B  
FILING DATE: 16-FEB-1998

```
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,447
; FILING DATE: 26-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: CLARK, JANET P.
; REGISTRATION NUMBER: 34,799
; REFERENCE/DOCKET NUMBER: R0020B-REG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 852-3097
; TELEFAX: (650) 855-5322
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1988 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-024-020B-4

Query Match 13.1%; Score 85; DB 2; Length 1988;
Best Local Similarity 25.2%; Pred. No. 4.1;
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

Qy 17 QFDGWEISGFEKGKQDAGYVINLSKDTFIKPVFKIEEKKBEENKPTFDVSKKKDNPOVNH 76
Db 1001 EMNLIQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKANCINAH 1056
Qy 77 SOLNESHKRELDQREHSQKSDSTKVDTATVLDKNNISSKSTTNNPN 123
Db 1057 TGV-DIHRNGDFQKNGGTSGIGSSVEKYIIDEDHM---SFINNPN 1099

RESULT 34
US-09-425-043-4
; Sequence 4, Application US/09425043
; Patent No. 6335172
; GENERAL INFORMATION:
; APPLICANT: DELGADO, STEPHEN G.
; APPLICANT: DIETRICH, PAUL S.
; APPLICANT: FISH, LINDA M.
; APPLICANT: HERMAN, RONALD C.
; APPLICANT: SANGAMESWARAN, LAKSHMI
; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE
; TITLE OF INVENTION: SODIUM CHANNEL 1-SUBUNIT AND A SPLICE VARIANT THEREOF
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JANET PAULINE CLARK
; STREET: 3401 HILLVIEW AVENUE, MS A2-250
; CITY: PALO ALTO
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 94304-1397
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/425,043
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/024,020
; FILING DATE: 16-FEB-1998
; APPLICATION NUMBER: US 60/039,447
; FILING DATE: 26-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: CLARK, JANET P.
; REGISTRATION NUMBER: 34,799
; REFERENCE/DOCKET NUMBER: R0020B-REG
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 852-3097
```

```
; TELEFAX: (650) 855-5322
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1988 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-425-043-4

Query Match 13.1%; Score 85; DB 2; Length 1988;
Best Local Similarity 25.2%; Pred. No. 4.1;
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

Qy 17 QFDGWEISGFEKGKQDAGYVINLSKDTFIKPVFKIEEKKBEENKPTFDVSKKKDNPOVNH 76
Db 1001 EMNLIQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKANCINAH 1056
Qy 77 SOLNESHKRELDQREHSQKSDSTKVDTATVLDKNNISSKSTTNNPN 123
Db 1057 TGV-DIHRNGDFQKNGGTSGIGSSVEKYIIDEDHM---SFINNPN 1099

RESULT 35
US-10-172-502-10
; Sequence 10, Application US/10172502
; Patent No. 6841154
; GENERAL INFORMATION:
; APPLICANT: POSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-10

Query Match 13.0%; Score 84.5; DB 2; Length 654;
Best Local Similarity 28.6%; Pred. No. 1.1;
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;

Qy 31 DAGYVINL-SKDTFIKPVFKIEEKKBEENKPTFDV-----SKKKDNPOVNHSQLNESH 84
Db 450 DGQYHVRIVDKEAFTKANTDKSNKKEQQDNSAKKEATPATPSKPTSPVKESSQKQDSQK 509
Qy 85 KEDLQ-----REHSQKSDSTKVDT-ATVLDKNNISSKSTTNNPNK 124
Db 510 DDNKQLPSVEKENDASSESGDKTPTATKPTKGEVSSSTT--PTK 552

RESULT 36
US-09-248-796A-20276
; Sequence 20276, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20276
```



; LENGTH: 817

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-20276

Query Match

12.9%; Score 84; DB 2; Length 817;

Best Local Similarity

27.0%; Pred. No. 1.6;

Matches 24; Conservative 17; Mismatches 34; Indels 14; Gaps 4;

QY 33 GVINLSK-----DTFKVPFKIEEKEEENKPTFDVSKKDNPNQVNHSQLNESHKED 87

Db 312 GLTISQSLDKLASTVQPIILDIEGKAE--KRQIDIEKKQ-----KELELQQLHEK-- 362

QY 88 LQREHSQKSDTKDVTATVLDKNISSK 116

Db 363 AKKEEHEAKEKEKRDIEIAKLERNQNDK 391

RESULT 37

US-09-949-016-8508

; Sequence 8508, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8508

; LENGTH: 461

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-8508

Query Match

12.8%; Score 83.5; DB 2; Length 461;

Best Local Similarity

21.7%; Pred. No. 0.87;

Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4;

QY 30 KDAGVINLSKDTFKVPFKIEKK-----EENKPTFDVSKKDNPNQ 73

Db 26 RDSGLSQBEEDTFIEE--QQLEEKLLERERLHEEWLLREQKAQEFRIKKEKEAA 83

QY 74 VN-----HSQLNESHKEDLQREHSQKSDTKDVTATVLD--KNNISSKSTTNP 122

Db 84 KKWLEEQERKLUKEQWKEQORKEEERQKQKEKEBAVQKMLDQAEINDLENSTTQNP 143

RESULT 38

US-09-248-796A-15008

; Sequence 15008, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 15008

; LENGTH: 476

; TYPE: PRT

; ORGANISM: Candida albicans

US-09-248-796A-15008

Query Match

12.8%; Score 83.5; DB 2; Length 476;

Best Local Similarity

27.6%; Pred. No. 0.91;

Matches 27; Conservative 19; Mismatches 27; Indels 25; Gaps 3;

QY 38 LSKDTFKVPFKIEEKEEENKPTFDVSKKKNPQVNH-----SQL 79

Db 181 LKDWKAK---LKLEKQKKNDELTKDLFKKQDTPASNPFGGNSNPFGLNPFSSK 237

QY 80 NESHKEDLQREHSQKSDTKDVTATVLDKNNISSKS 117

Db 238 PEEKEBEKEKEKETSYSKYADVAS----KNAPKPKS 271

RESULT 39

US-09-949-016-10508

; Sequence 10508, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10508

; LENGTH: 278

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-10508

Query Match

12.7%; Score 83; DB 2; Length 278;

Best Local Similarity

25.4%; Pred. No. 0.51;

Matches 30; Conservative 25; Mismatches 51; Indels 12; Gaps 4;

QY 8 YKGELEKGYQFDGWEISGPGKKDAGY--VINLSKDTFKVPFKIEEKEEENKPTFDV 65

Db 59 WEGEDEDVKNWDDDDDEKKEBAFVKPEVKISEK---KKIAEKIKKEKQKQKQBEI 115

QY 66 SKKDNFQ-----VNHSQLNESHKEDLQREHSQKSDTKDVTATV--LDKNNISSK 116

Db 116 KRLLEPEEPKULTPEEQSLADKLKLKQLQESDLELAKETFGVNNVATGIDAMPSSR 173

RESULT 40

US-09-248-796A-21334

; Sequence 21334, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstein et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA

; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 21334

```
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-21334

Query Match      12.7%; Score 82.5; DB 2; Length 243;
Best Local Similarity 30.8%; Pred. No. 0.48;
Matches 24; Conservative 16; Mismatches 27; Indels 11; Gaps 4;

QY 52 EKKEEENKPTDVS---KKQDNPOVNHSQLNESHKEDLQREHHSQKSDSTKDVATVL 108
   ||::: : ||| ||| : : ||| : : ||| : : ||| : : ||| : : ||| :
Db 71 EEEQQQQQIEDVSGYISPFDPNPIHSGK----HKRHLKND-SISNSSNKEITIDAI 125
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : ||| : : |||

QY 109 DK--NNISSKSTNNPN 123
   | |||| |||
Db 126 IKRQNNISSNDNNNNN 143
   |||||
```

Search completed: April 24, 2006, 15:03:35  
Job time : 21.2796 secs

GenCore version 5.1.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 77.8121 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_650\_773

Perfect score: 651

Sequence: 1 EDPLPVYKGLKGYQFDG.....ATVLDKNNISSKSTNNPNK 124

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_21:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*  
9: Geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	100.0	773	4	AAB48343
2	651	100.0	2120	3	AA81710 Streptococcus
3	651	100.0	2140	6	ABU01020 S. pneumo
4	651	100.0	2140	6	ABU45746 Protein e
5	651	100.0	2140	8	ADM92113 S. pneumo
6	651	100.0	2140	8	ADT50099 S. pneumo
7	648	99.5	637	8	ADR94534 Novel S.
8	648	99.5	637	9	AEA58404 Streptococcus
9	648	99.5	2138	8	ADK48759 Streptococcus
10	615	94.5	117	2	AAW55096 Streptococcus
11	615	94.5	117	5	ABP54590 Streptococcus
12	615	94.5	117	7	ADC45149 Streptococcus
13	106	16.3	188	9	ADZ79639 P. falcip
14	106	16.3	354	9	ADZ72253 Plasmodium
15	101.5	15.6	564	4	ABB61977 Drosophila
16	101	15.5	169	9	ADZ79634 P. falcip
17	101	15.5	647	9	ADZ79635 P. falcip
18	101	15.5	651	8	ADO19012 Amino acid
19	101	15.5	651	8	ADO19010 P. falcip
20	98	15.1	665	3	AB18278 Plasmodium
21	98	15.1	665	7	ABO23606 Plasmodium
22	96.5	14.8	707	6	ABU25018 Protein e
23	92.5	14.2	2468	6	ABR64281 Angiogene
24	92.5	14.2	2468	7	AD662723 Human Pro

25	92.5	14.2	2468	7	AD662719 Human Pro
26	92.5	14.2	2468	7	AD662727 Human Pro
27	92.5	14.2	2468	7	AD662715 Human Pro
28	92.5	14.2	2468	8	ADL12997 Human ste
29	92.5	14.2	2468	8	ADN05260 Antipsori
30	92.5	14.2	2468	8	ADR14614 Human NF-
31	92.5	14.2	2519	4	ASG16636 Novel hum
32	92.5	14.2	2527	8	ADN04561 Antipsori
33	91	14.0	470	8	ADT56185 Plant pol
34	91	14.0	484	3	AAQ47777 Arabidops
35	90.5	13.9	639	9	ADW88474 Staphyloc
36	89.5	13.7	511	2	AAV35091 Chlamydia
37	89.5	13.7	645	9	ADW88441 Staphyloc
38	88	13.5	225	5	ABP73992 Candida a
39	88	13.5	258	4	AA894584 Human pro
40	88	13.5	815	7	ADP28113 Murine gl
41	88	13.5	817	4	AAW79318 Human pro
42	88	13.5	817	4	AAW79319 Human pro
43	88	13.5	1980	7	ADH78600 Human bod
44	87.5	13.4	903	6	ABU24404 Protein e
45	87.5	13.4	1702	6	ADA09346 Haemophil

#### ALIGNMENTS

##### RESULT 1

AAB48343

ID AAB48343 standard; protein; 773 AA.

XX AAB48343;

AC AAB48343;

XX AAB48343;

DT 20-APR-2001 (first entry)

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

XX AAB48343;

CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Sp130 polypeptide  
XX  
SQ Sequence 773 AA;

Query Match 100.0%; Score 651; DB 4; Length 773;  
Best Local Similarity 100.0%; Pred. No. 1.3e-58;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKQDAGYVINLSKDTFIKPVFKKIEEKKEENK 60  
|||  
DB 650 EDFILPVYKGELEKGYQFDGWEISGPEGKQDAGYVINLSKDTFIKPVFKKIEEKKEENK 709  
|||

QY 61 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 120  
|||  
DB 710 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 769  
|||

QY 121 NPNK 124  
|||  
DB 770 NPNK 773  
|||

RESULT 2  
AAY81710  
ID AAY81710 standard; protein; 2120 AA.  
XX  
AC AAY81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PF 27-JUL-1999; 99WO-GB002452.  
XX  
PR 27-JUL-1998; 98GB-00016336.  
PR 19-MAR-1999; 99US-0125329P.  
XX  
PA (MICK-) MICROBIAL TECHNIQS LTD.  
XX  
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
DR WPI; 2000-195301/17.  
DR N-PSDB; AA291806.  
XX  
PT Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections.  
XX  
PS Claim 2; Page 41-42; 76pp; English.  
XX  
CC This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis  
XX  
SQ Sequence 2120 AA;

Query Match 100.0%; Score 651; DB 3; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 5e-58;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGPEGKQDAGYVINLSKDTFIKPVFKKIEEKKEENK 60  
|||  
DB 1963 EDFILPVYKGELEKGYQFDGWEISGPEGKQDAGYVINLSKDTFIKPVFKKIEEKKEENK 2022  
|||

QY 61 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 120  
|||  
DB 2023 PTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 2082  
|||

QY 121 NPNK 124  
|||  
DB 2083 NPNK 2086  
|||

RESULT 3  
ABU01020  
ID ABU01020 standard; protein; 2140 AA.  
XX  
AC ABU01020;  
XX  
DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX  
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KW gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
FN WO200277021-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-IB002163.  
XX  
PR 27-MAR-2001; 2001GB-00007658.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Masignani V, Tettelin H, Fraser C;  
XX  
DR WPI; 2003-040579/03.  
DR N-PSDB; ABX06302.  
XX  
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
PS Claim 1; SEQ ID NO 1180; 56pp; English.  
XX  
CC The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC ABS56454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target

CC sequence contained within a Streptococcus nucleic acid sequence, where  
CC the first primer is substantially complementary to the target sequence, and  
CC the second primer is substantially complementary to the complement of  
CC the target sequence, and where the parts of the primers having  
CC substantial complementarity define the termini of the target sequence to  
CC be amplified, assay comprising contacting a test compound with the  
CC protein, and determining whether the test compound binds to the protein  
CC and a Streptococcus pneumoniae bacterium, where one or more genes  
CC encoding the proteins has been rendered inactive. The proteins, nucleic  
CC acid molecules, antibody and compositions are useful as medicaments for  
CC treating or preventing a disease or infection due to streptococcus  
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
CC media or ear infection. They are also useful in developing vaccines,  
CC diagnostics and antibiotics. The methods are useful for identifying  
CC immunodominant proteins. The present sequence is one of the 2469 proteins  
CC expressed by the identified coding regions from the genomic sequence.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
CC standardise OS field)  
XX  
SQ Sequence 2140 AA;

Query Match 100.0%; Score 651; DB 6; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 5e-58;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKKEENK 60  
DB 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKKEENK 2042  
QY 61 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN 120  
DB 2043 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN 2102  
QY 121 NPNK 124  
DB 2103 NPNK 2106

RESULT 4  
ABU45746  
ID ABU45746 standard; protein; 2140 AA.  
XX  
AC ABU45746;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #31273.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Streptococcus pneumoniae.  
XX  
FN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PP 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
XX  
PR 06-SEP-2001; 2001US-00948993.  
XX  
PR 25-OCT-2001; 2001US-0342923P.  
XX  
PR 08-FEB-2002; 2002US-00072851.  
XX  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlseen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX N-PSDB; ACA49616.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 73670; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 2140 AA;

Query Match 100.0%; Score 651; DB 6; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 5e-58;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKKEENK 60  
DB 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKVPFKIEKKKEENK 2042  
QY 61 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN 120  
DB 2043 PTFDVSKKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN 2102  
QY 121 NPNK 124  
DB 2103 NPNK 2106

RESULT 5  
ADM92113  
ID ADM92113 standard; protein; 2140 AA.  
XX  
AC ADM92113;  
XX  
DT 03-JUN-2004 (first entry)  
XX  
XX S pneumoniae antigenic protein sequence SeqID310.  
XX  
XX antibacterial; gene therapy; Streptococcus pneumoniae infection;  
XX antigenic.  
XX  
XX Streptococcus pneumoniae.  
XX  
OS

PN WO2004020609-A2.  
 XX 11-MAR-2004.  
 XX 02-SEP-2003; 2003WO-US027401.  
 XX 30-AUG-2002; 2002US-0407082P.  
 XX (TUFT ) UNIV TUFTS.  
 XX Camilli A, Hava DL;  
 XX WPI; 2004-239189/22.  
 XX N-PSDB; ADM91876.  
 XX New Streptococcus pneumoniae nucleic acid molecules, useful for  
 PT diagnosing, treating and preventing active infections of Streptococcus  
 PT pneumoniae.  
 XX Claim 27; SEQ ID NO 310; 123pp; English.  
 XX This invention relates to novel isolated Streptococcus pneumoniae nucleic  
 CC acid molecules and the antigenic polypeptides encoded by them. The  
 CC invention may be useful for the production of compounds with an  
 CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
 CC compositions and methods disclosed are useful for treating Streptococcus  
 CC pneumoniae infection. The present sequence is that of an S pneumoniae  
 CC protein of the invention.  
 XX Sequence 2140 AA;  
 SQ  
 Query Match 100.0%; Score 651; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 5e-58;  
 Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENK 60  
 DB 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENK 2042  
 QY 61 PTFDVSKKDNQPNVNSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 120  
 DB 2043 PTFDVSKKDNQPNVNSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 2102  
 QY 121 NPNK 124  
 DB 2103 NPNK 2106  
 RESULT 6  
 ADT50099  
 ID ADT50099 standard; protein; 2140 AA.  
 XX AC ADT50099;  
 XX 13-JAN-2005 (first entry)  
 XX S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
 XX hyperimmune serum reactive antigen; antibacterial; vaccine;  
 KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
 KW sepsis; meningitis.  
 XX Streptococcus pneumoniae TIGR4.  
 XX WO2004092209-A2.  
 XX 28-OCT-2004.  
 XX 15-APR-2004; 2004WO-EP003984.  
 XX 15-APR-2003; 2003EP-00450087.  
 XX (INTE-) INTERCELL AG.

XX Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;  
 XX WPI; 2004-758335/74.  
 XX N-PSDB; ADT49955.  
 XX New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
 PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
 PT treating S. pneumoniae infections.  
 XX Disclosure; SEQ ID NO 177; 191pp; English.  
 XX This invention relates to novel nucleic acids encoding hyperimmune serum  
 CC reactive antigens, or fragments derived thereof. Specifically, it refers  
 CC to antigens selected from peptides and serum reactive epitopes that can  
 CC be used in pharmaceutical compositions that exhibit antibacterial  
 CC activity. The present invention describes a composition (including the  
 CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
 CC that is useful for manufacturing a medicament such as a vaccine, which  
 CC can be used to treat or prevent bacterial infections, particularly S.  
 CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
 CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
 CC be used for isolating, purifying and/ or identifying an interaction  
 CC partner of the hyperimmune serum reactive antigen, as well as for  
 CC manufacturing a functional nucleic acid selected from aptamers and  
 CC spiegelmers or for manufacturing a functional ribonucleic acid selected  
 CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
 CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
 CC of the invention.  
 XX Sequence 2140 AA;  
 SQ  
 Query Match 100.0%; Score 651; DB 8; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 5e-58;  
 Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENK 60  
 DB 1983 EDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEEKEENK 2042  
 QY 61 PTFDVSKKDNQPNVNSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 120  
 DB 2043 PTFDVSKKDNQPNVNSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTTN 2102  
 QY 121 NPNK 124  
 DB 2103 NPNK 2106  
 RESULT 7  
 ADT94534  
 ID ADT94534 standard; protein; 637 AA.  
 XX AC ADT94534;  
 XX 16-DEC-2004 (first entry)  
 XX Novel S. pneumoniae protein sequence, SEQ ID 3169.  
 XX Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
 KW bacterial infection.  
 XX Streptococcus pneumoniae.  
 XX US6800744-B1.  
 XX 05-OCT-2004.  
 XX 30-JUN-1998; 98US-00107433.  
 XX 02-JUL-1997; 97US-0051553P.  
 XX 12-MAY-1998; 98US-0085131P.



```

XX AC ADK48759;
XX DT 20-MAY-2004 (first entry)
XX DE Streptococcus pneumoniae protein, Seq ID No 5274.
XX KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX OS Streptococcus pneumoniae.
XX PN US6699703-B1.
XX PD 02-MAR-2004.
XX PF 26-MAY-2000; 2000US-00583110.
XX PR 02-JUL-1997; 97US-0051553P.
XX PR 12-MAY-1998; 98US-0085131P.
XX PR 30-JUN-1998; 98US-00107433.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX DR N-PSDB; ADK46098.
XX PT New nucleic acid molecules and polypeptides useful for diagnosing,
XX PT preventing and treating pathological conditions resulting from bacterial
XX PT infection, e.g. Streptococcus pneumoniae infection, and in drug
XX PT screening.
XX PS Disclosure; SEQ ID NO 5274; 301pp; English.
XX CC The invention relates to isolated Streptococcus pneumoniae nucleic acids
XX CC and polypeptides. The nucleic acids and proteins are useful for
XX CC diagnosing, preventing and treating pathological conditions resulting
XX CC from bacterial infection, such as S. pneumoniae infection. These may also
XX CC be used for drug screening procedures. The present sequence represents a
XX CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
XX CC data for this patent did not appear in the printed specification but was
XX CC obtained in electronic format directly from USPTO at
XX CC seqdata.uspto.gov/sequence.html.
XX SQ Sequence 2138 AA;

Query Match 99.5%; Score 648; DB 8; Length 2138;
Best Local Similarity 99.2%; Pred. No. 1e-57;
Matches 123; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 EDFILPVYKGELEKGYQFDGWEISGFEGKDAAGYVINLSKDTFIKPVFKIEEKEEENK 60
DB 1981 EDFILPVYKGELEKGYQFDGWEISGFEGKDAAGYVINLSKDTFIKPVFKIEEKEEENK 2040
QY 61 PTFDYSKKQNPQVNHSQLNESHKEDLQREHESQKSDSTKDVATVLDKNNISSKSTTN 120
DB 2041 PTFDYSKKQNPQVNHSQLNESHKEDLQREHESQKSDSTKDVATVLDKNNISSKSTTN 2100
QY 121 NFNK 124
DB 2101 NFNK 2104

RESULT 10
AAW55096
ID AAW55096 standard; protein; 117 AA.
XX AC AAW55096;
XX DT 02-OCT-1998 (first entry)
XX DE Streptococcus pneumoniae SP0043 protein.

```

```

XX KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX KW detection; pneumonia; otitis media; meningitis.
XX OS Streptococcus pneumoniae.
XX PN WO9818930-A2.
XX PD 07-MAY-1998.
XX PF 30-OCT-1997; 97WO-US019422.
XX PR 31-OCT-1996; 96US-0029960P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
XX DR WPI; 1998-272224/24.
XX DR N-PSDB; AAV27357.
XX PT Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
XX PT - or their epitope-containing fragments, useful in protective or
XX PT therapeutic vaccines, and for diagnosis.
XX PS Claim 11; Page 62; 118pp; English.
XX CC The present sequence represents a protein from Streptococcus pneumoniae.
XX CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX CC can be useful in vaccines for inducing protective antibodies against
XX CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX CC are used to detect Streptococcus infection (by usual hybridisation or
XX CC amplification methods), also for isolating Streptococcus genes or their
XX CC allelic variants. The protein can be used similarly to detect specific
XX CC antibodies in standard immunoassays, especially for diagnosing or
XX CC monitoring infections. Antibodies which bind the protein are used to
XX CC detect corresponding antigens, to purify the protein and for passive
XX CC immunisation (optionally coupled to a toxin). Vaccines are administered,
XX CC e.g. by injection, orally or through the skin, typically at 0.01-1000
XX CC (especially 10-300) mu g/ml per dose
XX SQ Sequence 117 AA;

Query Match 94.5%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.8e-56;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 YKGELEKGYQFDGWEISGFEGKDAAGYVINLSKDTFIKPVFKIEEKEEENKPTFDVSK 67
DB 1 YKGELEKGYQFDGWEISGFEGKDAAGYVINLSKDTFIKPVFKIEEKEEENKPTFDVSK 60
QY 68 KKDNPQVNHSQLNESHKEDLQREHESQKSDSTKDVATVLDKNNISSKSTTNPNK 124
DB 61 KKDNPQVNHSQLNESHKEDLQREHESQKSDSTKDVATVLDKNNISSKSTTNPNK 117

RESULT 11
ABP54590
ID ABP54590 standard; protein; 117 AA.
XX AC ABP54590;
XX DT 04-SEP-2002 (first entry)
XX DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.
XX KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX KW antibacterial; Streptococcal infection; detection.
XX OS Streptococcus pneumoniae.
XX PN US2002061545-A1.

```



```

XX PD 23-MAY-2002.
XX PF 22-JAN-2001; 2001US-00765272.
XX XX
XX PR 30-OCT-1997; 97US-00961083.
XX XX
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2002-479261/51.
XX DR N-PSDB; ABQ84825.
XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX PT and for preventing or attenuating disease caused by Streptococcus
XX PT infection.
XX PS Claim 11; Page 29; 70pp; English.
XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX CC pneumoniae antigens have antibacterial activity and can be used in
XX CC vaccines. The S. pneumoniae antigens can also be used to prevent or
XX CC attenuate a Streptococcal infection in an animal. The polynucleotides
XX CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX CC of S. pneumoniae ORFs (open reading frames) which are used in an example
XX CC from the present invention
XX SQ Sequence 117 AA;
XX Query Match 94.5%; Score 615; DB 5; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 5.8e-56;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 YKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 67
DB 1 YKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 60
QY 68 KKNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 124
DB 61 KKNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 117
RESULT 12
ADC45149
ID ADC45149 standard; protein; 117 AA.
XX AC ADC45149;
XX DT 18-DEC-2003 (first entry)
XX DE S. pneumoniae antigenic protein SP043.
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX OS Streptococcus pneumoniae.
XX XX
XX PN US6573082-B1.
XX PD 03-JUN-2003.
XX PF 28-MAR-2000; 2000US-00536784.
XX PR 31-OCT-1996; 96US-0029960P.

```

```

PR 30-OCT-1997; 97US-00961083.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2003-764574/72.
XX DR N-PSDB; ADC45148.
XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX PT useful for producing vaccines for prevention or attenuation of infection
XX PT by Streptococcus pneumoniae.
XX PS Example 1; SEQ ID NO 68; 58pp; English.
XX CC The invention relates to an isolated polynucleotide consisting of a
XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX CC antigens. Also included are making a recombinant vector by inserting the
XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at
XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae
XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX CC antigen nucleic acids are useful as probes for use in diagnostic methods
XX CC for detecting S. pneumoniae gene expression. The present sequence
XX CC represents an S. pneumoniae antigenic protein.
XX SQ Sequence 117 AA;
XX Query Match 94.5%; Score 615; DB 7; Length 117;
XX Best Local Similarity 100.0%; Pred. No. 5.8e-56;
XX Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 YKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 67
DB 1 YKGELEKGYQFDGWEISGFEGKKGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 60
QY 68 KKNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 124
DB 61 KKNPQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 117
RESULT 13
ADZ79639
ID ADZ79639 standard; protein; 188 AA.
XX AC ADZ79639;
XX DT 14-JUL-2005 (first entry)
XX DE P. falciparum merozoite surface protein 3, MSP3a to MSP3f fragment.
XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;
XX KW immunotherapy; malaria; antimalarial; vaccine.
XX OS Plasmodium falciparum.
XX XX
XX PN WO2005040206-A1.
XX PD 06-MAY-2005.
XX PF 22-OCT-2004; 2004WO-EP012910.
XX PR 24-OCT-2003; 2003US-00691672.
XX PA (INSP) INST PASTEUR.
XX PI Druilhe P;
XX XX
XX DR WPI; 2005-355821/36.

```

PT Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.

XX  
 PS Disclosure; SEQ ID NO 7; 79pp; English.

XX  
 CC The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also  
 CC described are: (i) a conjugate comprising the chimeric molecule of the  
 CC invention bound to a solid support, (ii) an immunogenic composition  
 CC comprising the chimeric molecule, the conjugate described above, or a  
 CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
 CC against malaria comprising the chimeric molecule, the conjugate described  
 CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
 CC association with a suitable vehicle, (iv) use of purified and/or  
 CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
 CC medicament against malaria, and (v) a medicament for passive  
 CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
 CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
 CC and MSP3 antigens are useful for the preparation of a vaccine composition  
 CC against malaria. This sequence represents Plasmodium falciparum MSP3a to  
 CC MSP3f fragment. Note: The present sequence given as SEQ ID No:7 in the  
 CC Sequence Listing is not mentioned elsewhere in the specification.

XX  
 SQ Sequence 188 AA;

Query Match 16.3%; Score 106; DB 9; Length 188;  
 Best Local Similarity 23.5%; Pred. No. 0.012;  
 Matches 32; Conservative 29; Mismatches 49; Indels 26; Gaps 5;  
 QY 7 VYKGELEKGYQFD--GWEISGF--EGKQDAG-----YVINLSKDTPIKPVFKKIEKKEE- 57  
 Db 15 VLKAKEASSYDYLGWFGGVPHEKKEENMLSHLYVSSKDKENISKENDDVLDKEBEEA 74  
 QY 58 -----ENKPTFDVSKKNPQVNHSQLNESHKEDLQREHSQKSDTKDYTA 105  
 Db 75 ETEEEELKEKNEETESEISEDEEEEBEKEEENDKKCKEKEQSNENNDDQKDMA 134  
 QY 106 TVLDKNNISSKSTTN 121  
 Db 135 -----QNLISKQNNN 145

RESULT 14

ADZ72253  
 ID ADZ72253 standard; protein; 354 AA.

XX  
 AC ADZ72253;

XX  
 DT 14-JUL-2005 (first entry)

XX  
 DE Plasmodium falciparum MSP-3-like protein, MSP-3-1 SEQ ID NO: 2.

XX  
 KW Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;  
 KW infection; merozoite surface protein 3-like protein; MSP-3-1; antigen.

XX  
 OS Plasmodium falciparum.

XX  
 PN EP1526178-A1.

XX  
 PD 27-APR-2005.

XX  
 PF 24-OCT-2003; 2003EP-00292673.

XX  
 PR 24-OCT-2003; 2003EP-00292673.

XX  
 PA (INSP ) INST PASTEUR.

XX

PI Druilhe P;

XX  
 DR WPI; 2005-323987/34.  
 DR N-PSDB; ADZ72252.

XX  
 PT Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
 PT falciparum, which encode proteins useful for preparing vaccine  
 PT compositions against malaria.

XX  
 PS Disclosure; SEQ ID NO 2; 137pp; English.

XX  
 CC The present invention relates to the protection against malaria. More  
 CC particularly, the invention pertains to a family of MSP-3 (merozoite  
 CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
 CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
 CC falciparum, highly conserved in P. falciparum strains, simultaneously  
 CC expressed in P. falciparum at the erythrocytic stages and encoding  
 CC proteins which have a Asn-Leu-Arg-Hen or Asn-Leu-Arg-Lys signature at  
 CC their N-terminal extremity and which are located at the merozoite  
 CC surface. The characterization of this gene family enables the definition  
 CC of immunogenic and vaccine compositions against P. falciparum. The  
 CC present sequence is the P. falciparum MSP-3-1 protein.

XX  
 SQ Sequence 354 AA;

Query Match 16.3%; Score 106; DB 9; Length 354;  
 Best Local Similarity 23.5%; Pred. No. 0.029;  
 Matches 32; Conservative 29; Mismatches 49; Indels 26; Gaps 5;  
 QY 7 VYKGELEKGYQFD--GWEISGF--EGKQDAG-----YVINLSKDTPIKPVFKKIEKKEE- 57  
 Db 181 VLKAKEASSYDYLGWFGGVPHEKKEENMLSHLYVSSKDKENISKENDDVLDKEBEEA 240  
 QY 58 -----ENKPTFDVSKKNPQVNHSQLNESHKEDLQREHSQKSDTKDYTA 105  
 Db 241 ETEEEELKEKNEETESEISEDEEEEBEKEEENDKKCKEKEQSNENNDDQKDMA 300  
 QY 106 TVLDKNNISSKSTTN 121  
 Db 301 -----QNLISKQNNN 311

RESULT 15

ABB61977

ID ABB61977 standard; protein; 564 AA.

XX  
 AC ABB61977;

XX  
 DT 26-MAR-2002 (first entry)

XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 12723.

XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX  
 OS Drosophila melanogaster.

XX  
 PN WO200171042-A2.

XX  
 PD 27-SEP-2001.

XX  
 PF 23-MAR-2001; 2001WO-US009231.

XX  
 PR 23-MAR-2000; 2000US-0191637P.

XX  
 PR 11-JUL-2000; 2000US-00614150.

XX  
 PA (PEKE ) PE CORP NY.

XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;

XX  
 DR WPI; 2001-656860/75.

XX  
 DR N-PSDB; ABL06080.











RESULT 25  
 ADE62719  
 ID ADE62719 standard; protein; 2468 AA.  
 XX  
 AC ADE62719;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human Protein AAA18904, SEQ ID NO 8652.  
 XX  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX  
 OS Homo sapiens.  
 XX  
 OS WO2003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 XX 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 XX (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 XX WPI; 2003-268312/26.  
 DR GENBANK; AAA18904.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 PT  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2468 AA;  
 Query Match 14.2%; Score 92.5; DB 7; Length 2468;  
 Best Local Similarity 31.0%; Pred. No. 10;  
 Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

OY 38 LSKDTFIKVPFKKIEKKKEENKPTFDVSKKONQVNHQSLSHRKDLRE-----E 92  
 Db 638 VKKETVKP-----EDKKEKEKPKIEVAKEDTPI---KKEKPKKEVKKEVKEIK 689  
 OY 93 HSQKSDSTKDV 103  
 Db 690 KEEKKEPKKEV 700  
 RESULT 26  
 ADE62727  
 ID ADE62727 standard; protein; 2468 AA.  
 XX  
 AC ADE62727;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human Protein AAA18904, SEQ ID NO 8660.  
 XX  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX  
 OS Homo sapiens.  
 XX  
 OS WO2003016475-A2.  
 PN  
 XX 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 XX 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 XX (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 XX WPI; 2003-268312/26.  
 DR GENBANK; AAA18904.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 PT  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 2468 AA;





CC specific binding; and detecting specific binding between each cDNA and at  
 CC least one molecule or compound. The molecules or compounds are regulatory  
 CC proteins. The combination is useful for preparing a composition for  
 CC treating liver disorders associated with steroid therapy, e.g., cirrhosis  
 CC or hepatitis. The present sequence represents a human protein which is  
 CC differentially expressed in steroid-induced C3A liver cells. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html.

XX Sequence 2468 AA;

Query Match 14.2%; Score 92.5; DB 8; Length 2468;  
 Best Local Similarity 31.0%; Pred. No. 10;  
 Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 38 LSKOTFIKPVFKIEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQRE-----E 92  
 Db 638 VKKETKVP-----EDKKEKPKKEVAKKEDKTPi---KKEKPKKEVKKEIK 689

QY 93 HSQKSDSTKDV 103

Db 690 KEEKKEPKKEV 700

RESULT 29

ADN05260

ID ADN05260 standard; protein; 2468 AA.

AC ADN05260;

XX 01-JUL-2004 (first entry)

DE Antipsoriatic protein sequence #805.

XX antipsoriatic; gene therapy; psoriasis; diagnosis.

XX Homo sapiens.

XX WO2004028479-A2.

PD 08-APR-2004.

XX 25-SEP-2003; 2003WO-US030907.

XX 25-SEP-2002; 2002US-0414006P.

XX (GETH ) GENENTECH INC.

XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
 PI Wu TD;

XX WPI; 2004-305105/28.

DR N-PSDB; ADN05259.

XX New PRO nucleic acid or polypeptide, useful for preparing a  
 PT pharmaceutical composition for diagnosing or treating psoriasis in a  
 PT mammal.

XX Claim 9; SEQ ID NO 1654; 3069pp; English.

XX The invention relates to novel polynucleotide and polypeptides for  
 CC treating psoriasis or a sequence having at least 80% identity to the  
 CC above sequences. The nucleic acid is useful for preparing a composition  
 CC for diagnosing or treating psoriasis in a mammal. This sequence  
 CC corresponds to one of the polypeptides of the invention.

XX Sequence 2468 AA;

Query Match 14.2%; Score 92.5; DB 8; Length 2468;  
 Best Local Similarity 31.0%; Pred. No. 10;  
 Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 38 LSKOTFIKPVFKIEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQRE-----E 92  
 Db 638 VKKETKVP-----EDKKEKPKKEVAKKEDKTPi---KKEKPKKEVKKEIK 689

QY 93 HSQKSDSTKDV 103

Db 690 KEEKKEPKKEV 700

RESULT 30

ADRI4614

ID ADRI4614 standard; protein; 2468 AA.

AC ADRI4614;

DT 21-OCT-2004 (first entry)

XX Human NF-kappaB pathway-associated protein SeqID615.

XX NF-kappaB pathway; antiinflammatory; cytostatic; hepatotropic; virucide;  
 KW antiarthritic; antirheumatic; gastrointestinal-Gen; antiasthmatic;  
 KW antiarteriosclerotic; immunomodulator; cerebroprotective; vasotropic;  
 KW immunosuppressive; vulnery; gene therapy; immune disorder;  
 KW inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;  
 KW hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;  
 KW hyper-IgM syndrome; hypohidrotic ectodermal dysplasia;  
 KW X-linked anhidrotic ectodermal dysplasia; immunodeficiency;  
 KW viral infection; HIV-1; HTLV-1; hepatitis B; hepatitis C; EBV; influenza;

XX viral replication; host cell survival; evasion of immune response;  
 KW rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;  
 KW atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAS;  
 KW autoimmune disorder; hyper immune activity;

XX aberrant acute phase response; hypercongenital condition; birth defect;

XX necrotic lesion; wound; organ transplant rejection;

XX aberrant signal transduction; proliferating disorder; cancer;

XX HIV propagation; human.

XX Homo sapiens.

XX WO2004065577-A2.

XX 05-AUG-2004.

XX 13-JAN-2004; 2004WO-US000798.

XX 14-JAN-2003; 2003US-0440068P.

XX 12-MAY-2003; 2003US-0469757P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Nadler SG, Neubauer MG, Feder JN, Carman J;

XX WPI; 2004-562168/54.

DR N-PSDB; ADRI4613.

XX New isolated polynucleotides and polypeptides associated with NF-kappaB  
 PT pathway, useful for diagnosing, treating, or preventing disorders or  
 PT diseases associated with NF-kappaB pathway.

XX Claim 6; SEQ ID NO 615; 237pp; English.

XX This invention relates to the novel association of protein sequences (and  
 CC the genes which encode them) to the NF-kappaB pathway. The invention may  
 CC be useful for the production of compounds with an antiinflammatory,  
 CC cytostatic, hepatotropic, virucide, antiarthritic, antiarteriosclerotic,  
 CC gastrointestinal-Gen, antiasthmatic, antiasthmatic, antineumatic,  
 CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or  
 CC vulnery activity or for gene therapy. The proteins and nucleotides are  
 CC useful for diagnosing, preventing, treating, or ameliorating conditions  
 CC or diseases associated with the NF-kappaB pathway. The condition is an  
 CC immune disorder, an inflammatory disorder, an inflammatory disorder  
 CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,  
 CC hepatic disorders, Hodgkin's lymphomas, haematopoietic tumours, hyper-IgM

CC syndromes, hypohidrotic ectodermal dysplasia, X-linked anhidrotic  
 CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,  
 CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell  
 CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory  
 CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick  
 CC syndrome, stroke, BAE, autoimmune disorders, disorders related to hyper  
 CC immune activity, disorders related to aberrant acute phase responses,  
 CC hypercongenital conditions, birth defects, necrotic lesions, wounds,  
 CC organ transplant rejection, conditions related to organ transplant  
 CC rejection, disorders related to aberrant signal transduction,  
 CC proliferating disorders, cancers and HIV propagation in cells infected  
 CC with other viruses. The present sequence is that of a human protein which  
 CC is subject to the novel association with the NF-kappaB pathway of the  
 CC invention. Note: This sequence does not appear in the specification but  
 CC was obtained by the indexer from Genbank.

XX Sequence 2468 AA;

Query Match 14.2%; Score 92.5; DB 8; Length 2468;  
 Best Local Similarity 31.0%; Pred. No. 10;  
 Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 38 LSKDTFKIPVKKIEKKKEENKPTFDVSKKDNQVNHSHKEDLQRE-----E 92

Db 638 VKKETKVKP-----EDKKEKEKPKKEVAKKEDKTPi---KKEKPKKEEVKKEIK 689

Qy 93 HSQKSDSTKDV 103

Db 690 KEEKPKKEV 700

RESULT 31

ABG16636

ID ABG16636 standard; protein; 2519 AA.

XX AC ABG16636;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #16627.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS80823.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX PS Claim 20; SEQ ID NO 46995; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 2519 AA;

Query Match 14.2%; Score 92.5; DB 4; Length 2519;  
 Best Local Similarity 31.0%; Pred. No. 10;  
 Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 38 LSKDTFKIPVKKIEKKKEENKPTFDVSKKDNQVNHSHKEDLQRE-----E 92

Db 689 VKKETKVKP-----EDKKEKEKPKKEVAKKEDKTPi---KKEKPKKEEVKKEIK 740

Qy 93 HSQKSDSTKDV 103

Db 741 KEEKPKKEV 751

RESULT 32

ADN04561

ID ADN04561 standard; protein; 2527 AA.

XX AC ADN04561;

XX DT 01-JUL-2004 (first entry)

XX DE Antipsoriatic protein sequence #471.

XX KW antipsoriatic; gene therapy; psoriasis; diagnosis.

XX OS Homo sapiens.

XX PN WO2004028479-A2.

XX PD 08-APR-2004.

XX PF 25-SEP-2003; 2003WO-US030907.

XX PR 25-SEP-2002; 2002US-0414006P.

XX PA (GETH ) GENENTECH INC.

XX PI Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;  
 PI Wu TD;

XX DR WPI; 2004-305105/28.

XX DR N-PSDB; ADN04560.

XX PT New PRO nucleic acid or polypeptide, useful for preparing a  
 PT pharmaceutical composition for diagnosing or treating psoriasis in a  
 PT mammal.

XX PS Claim 9; SEQ ID NO 955; 3069pp; English.

XX CC The invention relates to novel polynucleotide and polypeptides for  
 CC treating psoriasis or a sequence having at least 80% identity to the  
 CC above sequences. The nucleic acid is useful for preparing a composition  
 CC for diagnosing or treating psoriasis in a mammal. This sequence

CC corresponds to one of the polypeptides of the invention.  
 XX Sequence 2527 AA;  
 SQ

Query Match 14.2%; Score 92.5; DB 8; Length 2527;  
 Best Local Similarity 31.0%; Pred. No. 10;  
 Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 38 LSKDTFIKPVFKIEKKEENKPTFDYKKNQPNVNHSQLNESHKEDLQRE-----E 92  
 Db 697 VKKETKVP-----EDKKEKEKPKKEVAKKEDKTPI---KKBEKPKKEVKKEIK 748

QY 93 HSQKSDSTKDV 103  
 Db 749 KEEKPKKEV 759

RESULT 33  
 ADT56185  
 ID ADT56185 standard; protein; 470 AA.  
 XX  
 AC ADT56185;  
 XX  
 DT 13-JAN-2005 (first entry)  
 XX  
 DE Plant polypeptide, SEQ ID 6262.  
 XX

KW Plant; transgenic; cold tolerance; growth rate; drought tolerance;  
 KW disease resistance; galactomannan production; plant growth regulator;  
 KW heat tolerance; herbicide tolerance; lignin production;  
 KW extreme osmotic condition tolerance; pathogens resistance;  
 KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
 XX  
 OS Viridiplantae.  
 XX  
 PN US2004216190-A1.  
 XX  
 PD 28-OCT-2004.  
 XX  
 PF 18-DEC-2003; 2003US-00739930.  
 XX  
 PR 28-APR-2003; 2003US-00424599.  
 PR 28-APR-2003; 2003US-00425115.  
 XX  
 PA (KOVA/) KOVALIC D K.  
 XX  
 PI Kovalic DK;  
 XX  
 DR WPI; 2004-757369/74.  
 XX

PT New recombinant DNA constructs useful in the field of biochemistry and  
 PT genetics, and in particular for producing transgenic plants with improved  
 PT biological characteristics.  
 XX  
 PS Claim 2; SEQ ID NO 6262; 14pp; English.  
 XX

CC The invention relates a recombinant DNA construct comprising a  
 CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:  
 CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
 CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,  
 CC Arabidopsis, wheat and rape but the specification does not indicate which  
 CC sequences is derived from which organism. Also included is a method of  
 CC producing a plant having an improved property, comprising transforming a  
 CC plant with a recombinant DNA construct comprising a promoter region  
 CC functional in a plant cell operably joined to a polynucleotide encoding a  
 CC polypeptide associated with the property, and growing the transformed  
 CC plant. The property is selected from improving plant cold tolerance, for  
 CC manipulating growth rate in plant cells by modification of the cell cycle  
 CC pathway, for improving plant drought tolerance, for providing increased  
 CC resistance to plant disease, for galactomannan production, for production  
 CC of plant growth regulators, for improving plant heat tolerance, for  
 CC improving plant tolerance to herbicides, for increasing the rate of  
 CC homologous recombination in plants, for lignin production, for improving

CC plant tolerance to extreme osmotic conditions, for improving plant  
 CC tolerance to pathogens or pests, for yield improvement by modification of  
 CC photosynthesis, for modifying seed oil yield and/or content, for  
 CC modifying seed protein yield and/or content, for yield improvement by  
 CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
 CC and for yield improvement by providing improved plant growth and  
 CC development under at least one stress condition. The polynucleotide may  
 CC also encode a plant transcription factor. The methods and compositions of  
 CC the present invention are useful in the field of biochemistry and  
 CC genetics, in particular for producing transgenic plants with improved  
 CC biological characteristics such as increased yield, improved nitrogen  
 CC flow, increasing plant tolerance to cold or heat, improving plant  
 CC tolerance to extreme osmotic and drought conditions, and improving plant  
 CC tolerance to plant pests or pathogens. They can also be used in physical  
 CC arrays of molecules, plant breeding markers, computer-based storage and  
 CC analysis systems. The present sequence is one of the 5544 plant protein  
 CC sequences of the invention. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
 XX  
 SQ Sequence 470 AA;

Query Match 14.0%; Score 91; DB 8; Length 470;  
 Best Local Similarity 19.3%; Pred. No. 1.6;  
 Matches 29; Conservative 29; Mismatches 58; Indels 34; Gaps 4;

QY 8 YKGELEKGYFDGWEISGFE-----GKKDAGYVNLKSKDTPIKVPFKIEKKEEN 59  
 Db 100 YVQDLARRIRYDE-EATGSSQAKRIDHPNQKNGVITKAFENSPIEETSHRVDNKRINN 158

QY 60 KPTFDVSKKDN-----PQVHSQLNE-----SHRKEDLQREHS 94  
 Db 159 QKNFTAAKSSSENAVSRVSGADHKRAEVMGKPMENRDQVQTESAESKSHRKNVTKSEKP 218

QY 95 QKSDSTKDVATVLDKNNISSKSTNNPNK 124  
 Db 219 RDQEGVKKTEAKDKRNKKEKTEKTSINK 248

RESULT 34  
 AAG47777  
 ID AAG47777 standard; protein; 484 AA.  
 XX  
 AC AAG47777;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.  
 XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-00301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825P.  
 PR 05-MAR-1999; 99US-0123180P.  
 PR 09-MAR-1999; 99US-0123548P.  
 PR 23-MAR-1999; 99US-0125788P.  
 PR 25-MAR-1999; 99US-0126264P.  
 PR 29-MAR-1999; 99US-0126785P.  
 PR 01-APR-1999; 99US-0127482P.  
 PR 08-APR-1999; 99US-0128234P.  
 PR 16-APR-1999; 99US-0128714P.  
 PR 19-APR-1999; 99US-0129845P.  
 PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140821P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 09-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.

[illegible]

AC	AA135091;	
XX		
DT	17-OCT-2003 (revised)	
DT	13-SEP-1999 (first entry)	
XX		
XX		
DE	Chlamydia pneumoniae transmembrane protein sequence.	
XX		
XX		
KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;	
KW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis; vaccine;	
KW	neutralising epitope.	
XX		
OS	Chlamydophila pneumoniae.	
XX		
XX		
FN	WO9927105-A2.	
XX		
XX		
PD	03-JUN-1999.	
XX		
XX		
PP	20-NOV-1998; 98WO-IB001890.	
XX		
PR	21-NOV-1997; 97FR-00014673.	
PR	04-NOV-1998; 98US-0107078P.	
XX		
PA	(GEST ) GENSET.	
XX		
PI	Griffais R;	
XX		
DR	WPI; 1999-357842/30.	
XX		
PT	Genome sequence of Chlamydia pneumoniae.	
XX		
PS	Page 975-976; Disclosure; 1912pp; English.	
XX		
CC	AA134584-Y35879 represent the proteins encoded by all the open reading	

CC frames in the complete genome (see AAX91990) of *Chlamydia pneumoniae*. C.  
 CC pneumoniae causes respiratory disease such as pneumonia and bronchitis  
 CC and is thought to be a contributing factor in heart disease, sarcoidosis,  
 CC sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The  
 CC polypeptides encoded by the open reading frames of the C. pneumoniae  
 CC genome (see AAY34584-Y35879) can be used in immunogenic compositions as  
 CC vaccines. Vectors containing C. pneumoniae nucleotide sequences can also  
 CC be used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae. (Updated on 17  
 CC -OCT-2003 to standardise OS field)  
 XX  
 XX Sequence 511 AA;

Query Match 13.7%; Score 89.5; DB 2; Length 511;  
 Best Local Similarity 24.5%; Pred. No. 2.5;  
 Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
 QY 44 IKPVFKTIEKKKEENKPTFD-----VSKKDNQPNVHNSQLNESHK 85  
 DB 95 VGVFKTTPQAPFVSSPLPSHVHOGQRLPGLEGRDRIRKRSNPADLGKMKRSYSD 154  
 QY 86 EDLQREHSQKSDSTKDTATVLDKNNISSKSTT 119  
 DB 155 GDLDRVGHDSNEDSTEDSRS-----EGGEPSSKSS 185

RESULT 37  
 ADW88441  
 ID ADW88441 standard; protein; 645 AA.  
 AC ADW88441;

XX 21-APR-2005 (first entry)  
 DT Staphylococcus aureus hybrid polypeptide 0657nHybrid3.  
 DE  
 XX ORF0657n; vaccine; antibacterial; protein engineering;  
 KW Staphylococcus aureus infection; mutin.  
 KW  
 XX Staphylococcus aureus.  
 OS Synthetic.  
 OS  
 XX WO2005009378-A2.

XX  
 XX  
 XX PD 03-FEB-2005.  
 XX PF 22-JUL-2004; 2004WO-US023522.  
 XX PR 24-JUL-2003; 2003US-0489840P.  
 XX PA (MERI ) MERCK & CO INC.  
 XX PI Anderson AS, Kuklin N, Jansen KU;  
 XX  
 XX WPI; 2005-123069/13.

DR Novel hybrid polypeptide immunogen comprising modified ORF0657n sequence,  
 PT useful for inducing protective immune response in humans against  
 PT Staphylococcus aureus infection.  
 PS  
 PS Claim 7; SEQ ID NO 10; 84pp; English.

XX The present sequence is that of a *Staphylococcus aureus* protein ORF0657n  
 CC hybrid polypeptide denoted 0657nHybrid3. This is an example of claimed  
 CC hybrid polypeptide immunogens of the invention ADW8439-ADW8474 that  
 CC comprise a modified *S. aureus* ORF0657n sequence ADW8433-ADW8438  
 CC containing amino acid substitutions that increase sequence similarity to  
 CC ORF0190 ADW8432. The hybrid polypeptides contain one or more epitopes  
 CC for ORF0657n and ORF0190. They were designed by taking into account the  
 CC similarity and differences between native ORF0657n and ORF0190 protein  
 CC sequences. The invention also provides nucleic acids encoding these  
 CC hybrid polypeptides, and a method for evaluating the ability of an  
 CC immunogen to produce a protective immune response against *Staphylococcus*

CC infection using an animal (mouse or rat) model. The hybrid polypeptides  
 CC having therapeutic and diagnostic applications, such as being used to  
 CC provide protective immunity against *S. aureus* infection, being used to  
 CC generate antibodies to detect the presence of *S. aureus*, and being used  
 CC to generate therapeutic antibodies that target *S. aureus*.  
 XX  
 XX Sequence 645 AA;

Query Match 13.7%; Score 89.5; DB 9; Length 645;  
 Best Local Similarity 29.4%; Pred. No. 3.4;  
 Matches 32; Conservative 19; Mismatches 43; Indels 15; Gaps 5;  
 QY 26 PEGKKDAGYVNLKDTFKPVFKIIEKKKEENKPTFDV-----SKKDNQPNVHNSQLN 80  
 DB 440 YEGQY---HVRIIDKDAFTKANTDKSNKKEQDASAKKEATPATPSKFTPSFVEKESQK 496  
 QY 81 ESHREKEDLQ---REHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 124  
 DB 497 DSQKDDNKQLPSVEKENDASSESGDKTPTATKPTKGEVSSSTT--PTK 543

RESULT 38  
 ABP73992  
 ID ABP73992 standard; protein; 225 AA.  
 AC ABP73992;

XX 30-JAN-2003 (first entry)  
 DT Candida albicans essential protein SEQ ID NO 7829.  
 DE  
 XX Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;  
 KW signal transduction; DNA replication; cell division; growth;  
 KW proliferation; Candida albicans; fungicide; antifungal.  
 XX  
 OS Candida albicans.

XX WO200253728-A2.  
 XX 11-JUL-2002.  
 XX PF 26-DEC-2001; 2001WO-US049486.  
 XX PR 29-DEC-2000; 2000US-0259128P.  
 XX PR 20-FEB-2001; 2001US-00792024.  
 XX PR 22-AUG-2001; 2001US-0314050P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
 XX WPI; 2002-566694/60.  
 XX DR N-PSDB; ABZ32542.  
 XX

CC Constructing strains for identifying gene products as effective targets  
 CC for therapeutic intervention, by inactivating in the strain one allele of  
 CC a gene and placing other allele of the gene under conditional expression.  
 PS  
 PS Claim 44; SEQ ID NO 7829; 167pp + Sequence Listing; English.

XX The invention relates to constructing (M1) a strain of diploid fungal  
 CC cells in which both alleles of a gene are modified, comprising modifying  
 CC one allele by insertion or replacement by a cassette having an  
 CC expressible selectable marker and modifying other allele by  
 CC recombination, of a promoter replacement fragment with a heterologous  
 CC promoter, so that expression of the second allele is regulated by the  
 CC promoter. (M1) is useful for constructing a strain of diploid fungal  
 CC cells in which both alleles of a gene are modified. The diploid fungal  
 CC cells having both alleles modified are useful for identifying a gene that  
 CC is essential to the survival or growth of a fungus, a gene that  
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
 CC that contributes to the resistance of a diploid fungus to an antifungal  
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus

CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: the sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office  
XX

XX SQ Sequence 225 AA;

Query Match 13.5%; Score 88; DB 5; Length 225;

Best Local Similarity 28.0%; Pred. No. 1.2;

Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;

QY 16 YQFDGWEISGREGKDGAVVNLSDTKTTPVFKK---IREKKEENKPTFDVSKKK-- 69

DB 80 YDDDDPEGFSSNGAKELNLSQAIEKWKQRDLIEBEREKLNKKKEEIIIRKAKS 139

QY 70 --DNPVHNSQLNESHKEDLQREH--SQKSDSTKDVATVLDKNN 112

DB 140 TIDDFYENVNSKRDNHQKELLSEQKFIKRDDFLK--RGLWDRVN 184

RESULT 39

AAB94584

ID AAB94584 standard; protein; 258 AA.

XX AAB94584;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:15383.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 15383; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
CC specification. The primer sets can be used in antisense therapy and in  
CC gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AA92446 to AA95893  
CC represent human amino acid sequences, and AAH13629 to AAH13632 represent  
CC oligonucleotides, all of which are used in the exemplification of the  
XX present invention  
XX SQ Sequence 258 AA;

Query Match 13.5%; Score 88; DB 4; Length 258;

Best Local Similarity 26.3%; Pred. No. 1.4;

Matches 31; Conservative 25; Mismatches 50; Indels 12; Gaps 4;

QY 8 YKGELEKGYQFDGWEISGREGKDGAGY--VINLSKDTFIKVPFKIIEKKKEENKPTFDV 65

DB 39 WEGEDEDEDVKNWDDDDDEKKEANVKEVISEK---KKAIEKIKKEKQKKRQEEI 95

QY 66 SKKKNPQ----VNHSQLNESHKEDLQREHSSQKSDSTKDVATV--LDKNNISSK 116

DB 96 KKRLEPEEPKPVLTPEQLADKLRLKQLQESDLELAKETFGVNVTVYGDAMPSSR 153

RESULT 40

ADP28113

ID ADP28113 standard; protein; 815 AA.

XX AC ADP28113;

XX 12-FEB-2004 (first entry)

XX Murine glycosyl-phosphatidyl-inositol-anchored protein homologue.

XX neuroprotective; nootropic; cerebroprotective; antiparkinsonian;

XX neurological; spinal cord injury; cranial; cerebral trauma; stroke;

XX Alzheimer's disease; anxiety; autism; Parkinson's; tardive dyskinesia;

XX paralysis; seizure; memory disorder; adiponectin; C1q domain;

XX complement C1q; speract receptor; Wilm's tumour; synapsin; annexin;

XX leupin; serpin; NGRH; PRO; leucine-rich repeat; scavenger; neural IgCAM;

XX Ig F3; somatotropin; prolactin; somatostatin;

XX chorionic somatomammotropin hormone; NGAL; mucollipin; peroxidase; otx1;

XX SAPAP; murine; mouse.

XX Mus musculus.

XX WO2003048326-A2.

XX 12-JUN-2003.

XX 02-DEC-2002; 2002WO-US038526.

XX 03-DEC-2001; 2001US-00005499.

XX (HYSE-) HYSEQ INC.

XX Ghosh M, Tang YT, Wang JR, Wang Z, Zhao QA, Xu C, Mulero JJ;  
PI Boyle BJ;

XX WPI; 2003-513756/48.

XX N-PSDB; ADF28753.

XX New polynucleotides and polypeptides, useful for useful for treating  
XX neurological conditions, e.g. spinal cord injury, cranial or cerebral



```
PT trauma, stroke, Alzheimer's disease, anxiety, autism, Parkinson's
XX disease, or paralysis.
XX
XX Claim 9; SEQ ID NO 23; 396pp; English.
CC The invention relates to a novel isolated polynucleotide comprising any
CC of the 87 sequences fully defined in the specification or its mature
CC protein-coding portion. The polynucleotide of the invention demonstrates
CC neuroprotective, nootropic, cerebroprotective and antiparkinsonian
CC activities whilst the polynucleotides, polypeptides and compounds may be
CC useful for treating neurological conditions including spinal cord injury,
CC cranial or cerebral trauma, stroke, Alzheimer's disease, anxiety, autism,
CC Parkinson's disease, tardive dyskinesia, paralysis, seizures or memory
CC disorders. The current sequence is that of the protein of the invention.
XX
XX SQ Sequence 815 AA;

      Query Match          13 5%; Score 88; DB 7; Length 815;
Best Local Similarity    27.3%; Pred. No. 6.7;
Matches   33; Conservative 21; Mismatches 55; Indels 12; Gaps 4

Qy    10 GELEKGYQDGWEISG--PEGKKDAGYVYNLSKDTFTKPVKFKIEEKKEENKPTFDVSK 67
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db    196 GBKTKGESFKSEASGEKHQEVSKPAVSLEQRKDSTKLRLTPELREQKKQBIKSCKPSQ 255

Qy     68 -KKDNPVNHSOLNTSHRKED-----LQREBESHOKSDSTKDVTATVLDDKNINSSKT 118
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    256 WKQDTPRSGAGVQEHHKKGQETPKLPWPQLQEQ-DPPKQTPIGSWTFSMQEQQNTTKSWT 314

Qy    119 T 119
       |
Db    315 T 315
```

Search completed: April 24, 2006, 14:50:23  
Job time : 81.8121 secs

**THIS PAGE BLANK (08PT0)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 81.3893 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTVSEDFILPVYKG.....ATVLDKNNISKSTTNPKNK 134

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	696	100.0	2119	Q9AHT5_STRPN	Q9aht5 streptococc
2	696	100.0	2140	Q97RY6_STRPN	Q97ry6 streptococc
3	693	99.6	2144	Q9S4M8_STRPN	Q9s4m8 streptococc
4	693	99.6	2144	Q8DQP7_STRG6	Q8dqp7 streptococc
5	125.5	18.0	300	Q4XU16_PLACH	Q4xui6 plasmodium
6	112	16.1	361	Q5P15_PLAFA	Q5p15 plasmodium
7	110	15.8	346	Q9U0G0_PLARE	Q9u0g0 plasmodium
8	110	15.8	379	Q25705_PLAFA	Q25705 plasmodium
9	109	15.7	3008	Q8I436_PLAF7	Q8i436 plasmodium
10	108.5	15.6	600	Q77355_PLAF7	Q77355 plasmodium
11	108.5	15.6	1038	Q90784_CHICK	Q90784 gallus gall
12	108	15.5	384	Q50VJ0_ENTHI	Q50vj0 entamoeba h
13	107	15.4	354	Q25995_PLAFA	Q25995 plasmodium
14	107	15.4	354	Q8I353_PLAF7	Q8i353 plasmodium
15	107	15.4	829	Q8I5F3_PLAF7	Q8i5f3 plasmodium
16	106	15.2	616	Q6BRW2_DEBHA	Q6brw2 debaryomyce
17	103.5	14.9	379	Q8U6C4_PLAFA	Q8u6c4 plasmodium
18	103.5	14.9	380	Q26019_PLAFA	Q26019 plasmodium
19	103	14.8	296	Q50LX8_ENTHI	Q50lx8 entamoeba h
20	102.5	14.7	379	Q25706_PLAFA	Q25706 plasmodium
21	102	14.7	540	Q54MT2_DICDI	Q54mt2 dictyosteli
22	101.5	14.6	382	Q9V7J0_DROME	Q9v7j0 drosophila
23	101.5	14.6	556	Q9V7I9_DROME	Q9v7i9 drosophila
24	101.5	14.6	785	Q9GQ82_DROME	Q9gq82 drosophila
25	101.5	14.6	954	Q6HNR0_BACHK	Q6hnr0 bacillus th
26	101	14.5	662	Q4YMU4_PLABE	Q4ymu4 plasmodium
27	100	14.4	329	Q9NFV9_PLAFA	Q9nf9v plasmodium
28	100	14.4	1011	Q4Y213_PLACH	Q4y213 plasmodium
29	100	14.4	1130	Q8I3Z4_PLAF7	Q8i3z4 plasmodium
30	99.5	14.3	374	Q5V9M0_PLAKN	Q5v9m0 plasmodium
31	99	14.2	211	P91488_CABEL	P91488 caenorhabdi

32	99	14.2	437	2	Q54K26_DICDI	Q54k26 dictyosteli
33	99	14.2	467	2	Q59PE2_CANAL	Q59pe2 candida alb
34	99	14.2	467	2	Q59PL2_CANAL	Q59pl2 candida alb
35	99	14.2	1028	2	Q4Z4Q1_PLABE	Q4z4q1 plasmodium
36	99	14.2	1859	2	Q8IC27_PLAF7	Q8ic27 plasmodium
37	99	14.2	2563	2	Q8I3A0_PLAF7	Q8i3a0 plasmodium
38	98.5	14.2	393	2	Q7RKU2_PLAYO	Q7rku2 plasmodium
39	98.5	14.2	674	2	Q7RLR7_PLAYO	Q7rlr7 plasmodium
40	98.5	14.2	827	2	Q5SG46_DICDI	Q5sg46 dictyosteli
41	98	14.1	157	2	Q9VQV0_DROME	Q9vgv0 drosophila
42	98	14.1	951	2	Q8I229_PLAF7	Q8i229 plasmodium
43	97.5	14.0	449	2	Q8IHW3_PLAF7	Q8ihw3 plasmodium
44	97.5	14.0	556	2	Q95S93_DROME	Q95s93 drosophila
45	97.5	14.0	1345	1	YH00_YEAST	P38800 saccharomyc

ALIGNMENTS

RESULT 1  
Q9AHT5\_STRPN  
ID Q9AHT5\_STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5;  
DT 01-JUN-2001 (TREMblrel. 17, Created)  
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prta;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
DOI=10.1128/JAI.69.3.1593-1598.2001;  
RA Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
Langermann S., Johnson S., Koenig S.;  
RA "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 28BT.  
DR MEROPS; S08.064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN\_1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN\_1.  
KW Cell wall; Protease.

```

FT NON TER 1 1
SQ SEQUENCE 2119 AA; 238227 MW; 517F9B7F6B960A6A CRC64;

Query Match 100.0%; Score 696; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 4.6e-43;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDKDAGVYVNLKDTFIKPVFKK 60
Db 1952 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDKDAGVYVNLKDTFIKPVFKK 2011

Qy 61 IEKKKEENKPTFDVSKKKDNPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDK 120
Db 2012 IEKKKEENKPTFDVSKKKDNPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDK 2071

Qy 121 NNISKSSTNNPNK 134
Db 2072 NNISKSSTNNPNK 2085

RESULT 2
Q37RY6 STRPN
ID Q97RY6_STRPN PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Serine protease, subtilase family.
GN OrderedLocusNames=SP0641;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AB007373; AAK74791.1; -; Genomic_DNA.
DR PIR; F95074; F95074.
DR HSP; P00782; 2SST.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.

```

```

DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00679; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240426 MW; F744AD8E2938E334 CRC64;

Query Match 100.0%; Score 696; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 4.7e-43;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDKDAGVYVNLKDTFIKPVFKK 60
Db 1973 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDKDAGVYVNLKDTFIKPVFKK 2032

Qy 61 IEKKKEENKPTFDVSKKKDNPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDK 120
Db 2033 IEKKKEENKPTFDVSKKKDNPQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDK 2092

Qy 121 NNISKSSTNNPNK 134
Db 2093 NNISKSSTNNPNK 2106

RESULT 3
Q9S4M8 STRPN
ID Q9S4M8_STRPN PRELIMINARY; PRT; 2144 AA.
AC Q9S4M8;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase precursor PrtA.
GN Name=prtA;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3.B;
RX MEDLINE=21585565; PubMed=11728722;
RA Bethge G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,
RA Zysk G.;
RT "The cell wall-associated serine proteinase PrtA: a highly conserved
RT virulence factor of Streptococcus pneumoniae."
RL FEMS Microbiol. Lett. 205:199-204(2001).
DR EMBL; AF127143; AAD48399.1; -; Genomic_DNA.
DR HSP; P00782; 2SST.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.

```

```
DR PROSITE, PS00138; SUBTILASE_SER; UNKNOWN_1.
KW Cell wall; Signal.
FT SIGNAL 1 19 Potential.
FT CHAIN 20 2144 PTA.
FT SEQUENCE 2144 AA; 240725 MW; 20525114707411331 CRC64;
Query Match 99.6%; Score 693; DB 2; Length 2144;
Best Local Similarity 99.3%; Pred. No. 7.8e-43;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSKDTFIKPVFKK 60
DB 1977 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSKDTFIKPVFKK 2036
QY 61 IEKKEEENKPTFDVSKKONPQVNHSQLNESHKEDIQREHSHQKSDSTKDVATVLDK 120
DB 2037 IEKKEEENKPTFDVSKKONPQVNHSQLNESHKEDIQREHSHQKSDSTKDVATVLDK 2096
QY 121 NNISKSTTNNPNK 134
DB 2097 NNISKSTTNNPNK 2110
RESULT 4
QBDQP7_STRR6
ID Q8DQP7_STRR6 PRELIMINARY; PRT; 2144 AA.
AC Q8DQP7;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN Name=prtA; OrderedLocNames=spr0561;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Strem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Mateuashima P.,
RA McLaren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peary R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE08434; AAK93365.1; -; Genomic_DNA.
DR FJ; A97942; A97942.
DR HSP; P00782; 2SPT.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042803; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR00209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00746; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;
Query Match 99.6%; Score 693; DB 2; Length 2144;
Best Local Similarity 99.3%; Pred. No. 7.8e-43;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSKDTFIKPVFKK 60
DB 1977 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSKDTFIKPVFKK 2036
QY 61 IEKKEEENKPTFDVSKKONPQVNHSQLNESHKEDIQREHSHQKSDSTKDVATVLDK 120
DB 2037 IEKKEEENKPTFDVSKKONPQVNHSQLNESHKEDIQREHSHQKSDSTKDVATVLDK 2096
QY 121 NNISKSTTNNPNK 134
DB 2097 NNISKSTTNNPNK 2110
RESULT 5
Q4XUI6_PLACH
ID Q4XUI6_PLACH PRELIMINARY; PRT; 300 AA.
AC Q4XUI6;
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Hypothetical protein (Fragment).
GN ORFNames=PC000286.03.0;
OS Plasmodium chabaudi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5825;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras S.M., Raine J.D., Carlton J.M., Koolij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Jense C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:92-96(2005).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAATJ01003049; CAH79425.1; -; Genomic_DNA.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR PRINTS; PR00320; GPROTEINERPT.
DR ProDom; PD0000018; WD40; 3.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS00678; WD_REPEATS_2; 4.
DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
FT NON_TER 1
SQ SEQUENCE 300 AA; 34469 MW; 8ED4E512AFB1945 CRC64;
Query Match 18.0%; Score 125.5; DB 2; Length 300;
Best Local Similarity 28.7%; Pred. No. 0.18;
Matches 33; Conservative 26; Mismatches 31; Indels 25; Gaps 6;
QY 15 LPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSKDTFIK-----PVFKIEEKKE 66
```

Db 164 LAVYRGHVPAYKI-AWSI-----DNNYVSCSDSTLKLWRINHLVPLKKENAE 215  
Qy 67 ---BENKTFPDVSKKQNDPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVL 118  
Db 216 QTKDQK-----NEQKENPQ-NNDQPNDEANSEKKKKCKEKNKDKTKNKIKTL 264

## RESULT 6

ID Q95P15\_PLAPA PRELIMINARY; PRT; 361 AA.  
AC Q95P15;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN-PVO;  
RX MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;  
RA Hiseida H., Saul A., Reece J.J., Kennedy M.C., Long C.A., Miller L.H.,  
RA Stowers A.W.;  
RT "Merozoite surface protein 3 and protection against malaria in Aotus  
RT nancymai monkeys";  
RL J. Infect. Dis. 185:657-664 (2002).  
DR EMBL; AY044180; AAK94780.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
KW Merozoite.  
FT CHAIN <1 >361 merozoite surface protein 3.  
FT NON\_TER 1 1  
FT NON\_TER 361 361  
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 16.1%; Score 112; DB 2; Length 361;  
Best Local Similarity 23.9%; Pred. No. 2.2; Mismatches 48; Indels 42; Gaps 6;  
Matches 37; Conservative 28;  
Qy 18 YKGELEKGYQ-----PD---GWEISGF--EGKQDAG-----YVI 46  
Db 165 YAGKVKDYERAKYQKAVLKAKEASSDYILGWFGGVPEHKKEENMLSHLYVS 224  
Qy 47 NLSKDTFTKPVFKKEEENKPTFDVSKKQNDPQVNHSQLNESHKEDLQREHSQK 106  
Db 225 SKOKENISKENDVDLDE-KEEAETETEEELKEKNEETETSEISEDEEEEEEKEE 283  
Qy 107 SDSTKDVATVLDKN-----NISKSTTN 131  
Db 284 NDKKKEQKEQSNENNNDQKQMEANLISKQNNN 318

## RESULT 7

ID Q9U0G0\_PLARE PRELIMINARY; PRT; 346 AA.  
AC Q9U0G0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
GN Name=msp3;  
OS Plasmodium reichenowi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5854;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;  
RA Okenu D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
RT Plasmodium reichenowi and Plasmodium falciparum";  
RL Mol. Biochem. Parasitol. 109:185-188 (2000).

DR EMBL; AJ252286; CAB65754.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
KW Merozoite.  
FT NON\_TER 1 1  
FT NON\_TER 346 346  
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAFAP010 CRC64;  
Query Match 15.8%; Score 110; DB 2; Length 346;  
Best Local Similarity 26.2%; Pred. No. 2.9; Mismatches 54; Indels 16; Gaps 5;  
Matches 34; Conservative 26;  
Qy 17 VYKGELEKGYQFD-GWEISGF--EGKQDAG-----YVINLSKDTFTKPVFKKEEKEE 68  
Db 184 VLKAKBASSYNYILGWFGGVPEHKKEENMLSHLYVSSKDKENISKENDVDLDE-KEE 242  
Qy 69 NKPTFDVSKKQNDPQVNHSQLNESHKEDLQREHSQKSDSTKDVATV-----LDKN 121  
Db 243 AEETGEQELKEKNEETETSEISEDEKEEENNDKKEQAKESQNDQKEDMEAQ 302  
Qy 122 NISKSTTN 131  
Db 303 NLISKQNNN 312

## RESULT 8

ID Q25705\_PLAPA PRELIMINARY; PRT; 379 AA.  
AC Q25705;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31 (1997).  
DR EMBL; U08851; AAC47831.1; -; Unassigned\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 15.8%; Score 110; DB 2; Length 379;  
Best Local Similarity 23.4%; Pred. No. 3.2; Mismatches 41; Indels 50; Gaps 7;  
Matches 37; Conservative 30;  
Qy 18 YKGELEKGYQ-----PD---GWEISGF--EGKQDAG-----YVI 46  
Db 185 YAGKVKDYERAKYQKAVLKAKEASSDYILGWFGGVPEHKKEENMLSHLYVS 244  
Qy 47 NLSKDTFTKPVFKKEEENKPTFDVSKKQNDPQVNHSQLN-----ESH 93  
Db 245 SKOKENISKENDVDLDE-KEEAETETEEELKEKNEETETSEISEDEEEEEEKEEENE 303  
Qy 94 RKEDLQREHSQKSDSTKDVATVLDKNISKSTTN 131  
Db 304 KKEQKEQESNENNNDQKQMEAE-----QLISKQNNN 336

## RESULT 9

ID Q8I436\_PLAF7 PRELIMINARY; PRT; 3008 AA.  
AC Q8I436;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein PFE0325w.

GN Name=PF80325w;  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 ON NCBI\_TaxID=36329;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3D7;  
 EX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Sulston J.E., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
 RL Nature 419:527-531(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3D7;  
 RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL929351; CAD51431.1; -: Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 3008 AA; 356025 MW; 60BCBBEB15C599B4 CRC64;  
 Query Match 15.7%; Score 109; DB 2; Length 3008;  
 Best Local Similarity 32.4%; Pred. No. 35;  
 Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;  
 QY 44 YVINLSK----DTFTKPKFKIEKKKEENKPTFDVSKKKNPQVNHQSQ---LNESHKKE 96  
 Db 2310 YDIELSKLEKGCASIGPFTD--EENKKEKN--EVNKEENKKEENKKEENKKE 2366  
 QY 97 DLQREH----SQKSDSTKVATVLDKRNISK-----STNNPNK 134  
 Db 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414  
 RESULT 10  
 ID O77355 PLAF7 PRELIMINARY; PRT; 600 AA.  
 AC O77355;  
 DT 01-NOV-1998 (TRENBLrel. 08, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DE Hypothetical protein MAL3P4.20.  
 GN Name=MAL3P4.20; Synonyms=PF0465C;  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 ON NCBI\_TaxID=36329;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;  
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,  
 RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
 RA Horrocks P., Jagels K., Jaseal B., Kyes S., Mclean J., Moule S.,  
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,  
 RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;  
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum."  
 RL Nature 400:532-538 (1999).

RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Sulston J.E., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
 RL Nature 419:527-531(2002).  
 RN [1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 DR EMBL; AL008970; CAAL5610.2; -: Genomic\_DNA.  
 DR PIR; T18467; T18467.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0006397; P:mRNA processing; IEA.  
 DR InterPro; IPR02483; PWI.  
 DR Pfam; PF01480; PWI; 1.  
 DR SMART; SM00311; PWI; 1.  
 KW Hypothetical protein; Lyase.  
 SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;  
 Query Match 15.6%; Score 108.5; DB 2; Length 600;  
 Best Local Similarity 29.3%; Pred. No. 6.8;  
 Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;  
 QY 7 IVSEEDILPVY-----KGELEKGYQDGEWISGFEKK-----DAGYVINLSKDTFKPV 57  
 Db 60 ILGFEDILYCYCISQLKQSKKK---DGBEDKYLNAKCLKINLTGFTGNKKSDFIBEL 116  
 QY 58 FKKI--BEKKBE-----ENKPTFDVSK--KKNPQVNHQSQNLN-----SHRK 95  
 Db 117 LELLNEKKKEEHTADTLNENK--TNDIKKVNENENENYNNENKDNKDKKHVSHQN 175  
 QY 96 E-----DLQREH-----SQKSDSTK-----DVTATVLDKRNISKSTTN 130  
 Db 176 EHNINNVNKKKEYTDTIQDKRKHKESLSQKSDSYKKRPKNKKTSTIER--SLSNKRYDE 234  
 QY 131 NPNK 134  
 Db 235 KTNK 238  
 RESULT 11  
 ID Q90784 CHICK PRELIMINARY; PRT; 1038 AA.  
 AC Q90784;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE Claustrin.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 ON NCBI\_TaxID=9031;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC TISSUE=Brain;  
 RX MEDLINE=94157526; PubMed=7906711;  
 RA Burg M.A., Cole G.J.;  
 RT "Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally related to MAP1B."  
 RL J. Neurobiol. 25:1-22(1994).





RESULT 16  
 Q6BRW2 DEBHA PRELIMINARY; PRT; 616 AA.  
 AC Q6BRW2;  
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Similar to CA4458|IPF8464 Candida albicans IPF8464 unknown function.  
 DE  
 GN OrderedLocusNames=DEHA0D14674g;  
 OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.  
 NCBI\_TaxID=4959;  
 RN  
 RX NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RP STRAIN=ATCC 36239 / CBS 767;  
 RX PubMed=15229592; DOI=10.1038/nature02579;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,  
 RA Goffard N., Frangoul L., Algile M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
 RA Boixame A., Boyer J., Cattolico L., Confaniolieri F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire R., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,  
 RA Swennen D., Tekala F., Wesolowski-Leuvel M., Westhof E., Wirth B.,  
 RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.-L.;  
 RL "Genome evolution in yeasts."  
 RT Nature 430:35-44(2004).  
 DR EMBL; CR382136; CAG87226.1; -: Genomic DNA.  
 DR GO; GO:0016303; F:kinase activity; IEA.  
 DR InterPro; IPR000749; ATP\_gua\_Ptrans.  
 KW Complete proteome.  
 SQ SEQUENCE 616 AA; 72143 MW; 884009B2B8B6C3CF CRC64;  
  
 Query Match 15.2%; Score 106; DB 2; Length 616;  
 Best Local Similarity 30.7%; Pred. No. 11;  
 Matches 43; Conservative 20; Mismatches 49; Indels 28; Gaps 9  
  
 QY 16 PVVKGLEKGYGFDGWEISGPEGKADAGYVNLISKDT-FIKPV----FKIIEKKGEENK 70  
 DB 87 PVLKGRRAKFKITLDTIT-----KDLN-DINFSDSSEHKFPETSKKEKTKTKYK 140  
  
 QY 71 PTFDVSKKK-----DNQP-VNHSQALNE---SHRKDLQREHSQK-----SDSTKQVT 114  
 DB 141 PLDIDIGKLERRIVADNPDEIENHSSESEIKQRKEKRQKQDKQRKLKANKQCESNNDST 200  
  
 QY 115 ATVLDK-NNISKSTNNPN 133  
 DB 201 TQSPBLKNNINEKITSNEPS 220  
  
 RESULT 17  
 Q9U6C4 PLAF4 PRELIMINARY; PRT; 379 AA.  
 AC Q9U6C4;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Polymorphic antigen.  
 GN Name=WSF-3;  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=5833;  
 RN  
 RX NUCLEOTIDE SEQUENCE.  
 RP STRAIN=PFCCJ/HN;  
 RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;  
 [1]

RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=9815743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;  
RA McColl D.J., Anders R.P.;  
RT "Conservation of structural motifs and antigenic diversity in the

QY 110 TKDVTATVLDKNNISKSSTTNPKN 134  
:  
::  
|||

DB 495 SKKEEPVKKEKSSKKEDKKEK 519

RESULT 22

Q9V7J0 DROME  
AC Q9V7J0 DROME PRELIMINARY; PRT; 382 AA.  
ID Q9V7J0 Q9G081;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
GN CG8421-BP, isoform B (Aspartyl beta-hydroxylase variant 2).  
DE Names=ASPH; ORFNames=CG8421;  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
[1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista A.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabrielian A.E., Garcia N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kaiml B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster."  
RT Science 287:2185-2195 (2000).  
[2]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
RA Fasel R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
RT melanogaster euchromatic genome sequence."  
RT Genome Biol. 3:RESEARCH0079-RESEARCH0079 (2002).

```

RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RN NUCLEOTIDE SEQUENCE.
RP Berkeley Drosophila Genome Project;
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirkas R., Smith E.,
Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RN NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
RN [7]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=20564328; PubMed=10956665; DOI=10.1074/jbc.M006753200;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved
isoform of Asph missing the catalytic domain share exons with
junctin.";
RL J. Biol. Chem. 275:39543-39554(2000).
DR EMBL; AF003808; AAF58063.2; -; Genomic_DNA.
DR EMBL; AF289494; AAG40807.1; -; mRNA.
DR Ensembl; CG8421; Drosophila melanogaster.
DR FlyBase; FBgn0034075; Asph.
DR FlyBase; FBgn0034075; CG8421.
SQ SEQUENCE 382 AA; 43287 MW; 60B5C03AEBFC6E8B CRC64;

Query Match 14.6%; Score 101.5; DB 2; Length 382;
Best Local Similarity 24.5%; Pred. No. 14;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 11 EDPLPVYGELEKGVQPGW-----ETSGEGKKDAGYV-----NLKDTFTK 55
Db 78 EDLDTPLSRRSK--VFDGWVDEHDEHDGHVQPSGEALDDHDEHDDHDEDEEE 135
QY 56 PVFKTEERKEENKPT-----FDVSKKKDPQVNHSQLNESHKREDLQREHSQKSDS 109
Db 136 PLTEELLEELEEEETPEDEPADEYEDEDEENNA--GENITAEDAEDEEDND 193

QY 110 TKDVTATVLDKNISKST 128
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 23
Q9V719 DROME
ID Q9V719 DROME PRELIMINARY; PRT; 556 AA.
AC Q9V719;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
```

```

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG8421-PD, isoform D (CG8421-pe, isoform e).
GN Name=Asph; ORFNames=CG8421;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hestis N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan J.A.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shier B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RN NUCLEOTIDE SEQUENCE.
RP MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RN NUCLEOTIDE SEQUENCE.
RP
```

RX MEDLINE=22426069; PubMed=12537572;  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celnik S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Berkeley Drosophila Genome Project;  
 RA Celnik S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hoskins R., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,  
 RA Yu C., Rubin G.;  
 RT "Drosophila melanogaster release 4 sequence.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP NUCLEOTIDE SEQUENCE.  
 RG FlyBase;  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
 CC O46085.EG:63B12.5; NbExp=1; IntAct=EBI-123244, EBI-151469;  
 DR EMBL; AB003808; AAF58064.2; -; Genomic\_DNA.  
 DR InterAct; Q9V719; -;  
 DR Ensembl; CG8421; Drosophila melanogaster.  
 DR FlyBase; FBgn0034075; Asph.  
 DR FlyBase; FBgn0034075; CG8421.  
 SQ SEQUENCE 556 AA; 63144 MW; B420980CDB6C357A CRC64;  
 Query Match 14.6%; Score 101.5; DB 2; Length 556;  
 Best Local Similarity 24.5%; Pred. No. 21;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 11 EDPLPVYKGELEKGYQFGW-----BISPEGKKDAGYVI-----NLKDPFIK 55  
 DB 78 EDLDTPLSESFRSK--VFDGWDVDEHGDHGVQPSGEALDDHDDHDDHDEDEE 135  
 QY 56 PVFKKIRKKEENKPT-----PDVSKKDNPNVNSQLNESHKRLQREHSOKSDS 109  
 DB 136 PLTELELELEEEETDEPADEYEYDEDEENNA--GENTTAEDAESEEDND 193  
 QY 110 TKDVTATVLDKNNISSKST 128  
 DB 194 EGTVEATVTEATTAT 212  
 RESULT 24  
 Q9G082 DROME PRELIMINARY; PRT; 785 AA.  
 AC Q9G082;  
 DT 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA, isoform A).  
 GN Name=Asph; ORFNames=CG8421;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI TaxID=7227;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=20564328; PubMed=10956665; DOI=10.1074/jbc.M006753200;  
 RA Dinichuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
 RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
 RA Friedman P.A.;  
 RT "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved  
 RT isoform of Asph missing the catalytic domain share exons with  
 RT junction.";  
 RL J. Biol. Chem. 275:39543-39554(2000).

RN NUCLEOTIDE SEQUENCE.  
 RP MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H.F., Agbayani A., An H.-J., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Bayraktaroglu L., Bayraktaroglu L., Beasley E.M.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Bokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celnik S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 RT melanogaster euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celnik S.E.;  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
 RN [5]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celnik S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a

```
RT systematic review.;
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celinker S., Carlson J., Wan K., Pfeiffer B., Fries E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirskas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF298493; AAG40806.1; -; mRNA.
DR EMBL; AE003808; AAM70947.1; -; Genomic DNA.
DR Ensembl; CG8421; Drosophila melanogaster.
DR FlyBase; FBgn0034075; Asph.
DR FlyBase; FBgn0034075; CG8421.
DR GO; GO:0030176; C:integral to endoplasmic reticulum membrane; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0004597; F:peptide-aspartate beta-dioxygenase activity; IEA.
DR GO; GO:0018193; F:peptidyl-amino acid modification; IEA.
DR InterPro; IPR007803; Asp Arg Hydrol.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001440; TPR-like helical.
DR InterPro; IPR001990; TPR-like helical.
DR Pfam; PF05118; Asp Arg Hydrol; 1.
DR PROSITE; PS02993; TPR REGION; 1.
DR PROSITE; PS00142; ZINC PROTEASE; UNKNOWN 1.
SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD6836F7F1 CRC64;

Query Match 14.6%; Score 101.5; DB 2; Length 785;
Best Local Similarity 24.5%; Pred.No. 30;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 11 EDFILPVYKGELEKGVQFDGW-----EISGEGKKDAGYVI-----NLSKDTIK 55
DB 78 EDLDTPLSESRFSK-VFDGWVDEHRDHDGHVQEPGSEALDDHDEHDDHDDHDEDEE 135
QY 56 PVFKIEEKKKEENKPT-----FDVSKKKNQPNVHSQLNESHKRDLPQREHSQKSDS 109
DB 136 PLTEELEELEREEPTREDEPADEVEYDEDEENNA--GENITAEADAEEREEEDND 193
QY 110 TKDVTATVLDKNNISKST 128
DB 194 EGTVEATVETATTEAT 212

RESULT 25
Q6HNR0 BACHK PRELIMINARY; PRT; 954 AA.
AC Q6HNR0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Possible internalin protein.
GN OrderedlocusNames=BT9727_0463;
OS Bacillus thuringiensis (subsp. konkukian).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group.
ON NCBI_TaxID=180856;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=97-27;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RT "Complete genome sequence of Bacillus thuringiensis 97-27.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017355; AAT63966.1; -; Genomic DNA.
DR GO; GO:0009986; C:cell surface; IEA.
```

```
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003885; LRR_cyst.
DR InterPro; IPR007092; LRR_SDS22.
DR InterPro; IPR006635; NEA_transpt.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF00560; LRR_1; 8.
DR Pfam; PF05031; NEAT; 1.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00365; LRR_SD22; 8.
DR SMART; SM00725; NEAT; 1.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS09078; NEAT; 1.
KW Complete proteome.
SQ SEQUENCE 954 AA; 108589 MW; 4F4CF8B44C9B355F CRC64;

Query Match 14.6%; Score 101.5; DB 2; Length 954;
Best Local Similarity 25.7%; Pred.No. 37;
Matches 43; Conservative 32; Mismatches 43; Indels 49; Gaps 9;

QY 11 EDFILPVY--KGEL-----EKGYQP--DGWEIS--GFEG--- 38
DB 727 ERIQPVVDLEGEIENIKUTSDGTNNGVKWTGPEKVKYKFDLOSDEISFNFGTVI 786
QY 39 -----KKDAGYVINLSKDTFIKPVFKIEEKKKEENKPTFDVSKKKNQPNVHSQLN 90
DB 787 QNIVEKEEKEPTKEVEESKEEKEPT--KEVEESKEEKEPTKEVEESKEEVEPTKEVE 845
QY 91 ESHRKEDLPQREHSQKSDSKDVKDTATVLDKN-----NLSKSTNNPN 133
DB 846 ES--KEEV--KEPTKEVEESKEEVAJEIIEKKEEINQSAVPQEQNVN 888

RESULT 26
Q4YMU4 PLABE PRELIMINARY; PRT; 662 AA.
ID Q4YMU4 PLABE PRELIMINARY;
AC Q4YMU4;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE RNA binding protein, putative.
GN ORFNames=PB001104.03.0;
OS Plasmodium birgei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=5821;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAT01003467; CAI00666.1; -; Genomic DNA.
DR InterPro; IPR012972; NLE.
DR InterPro; IPR01680; WD40.
DR Pfam; PF08154; NLE; 1.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR PRODOM; PD000018; WD40; 4.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 662 AA; 75536 MW; E7C8543AF5E59124 CRC64;
```









```

OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Kooij T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Jase C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.,
RA "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CA0100607; CAH94732.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1028 AA; 120493 MW; 866A9FFPCC427612 CRC64;
Query Match 14.2%; Score 99; DB 2; Length 1028;
Best Local Similarity 25.0%; Pred. No. 62;
Matches 41; Conservative 22; Mismatches 53; Indels 48; Gaps 6;
QY 10 BEDFILPVYKGELEGYQFDGWEISGFGKKDAGYVINSKDTFTKPVFKK----- 60
Db 566 DDDYYDYDEYSEYKGEK-----KKVFEGKK-----NLKNNKKEKLNKGKNNKINE 616
QY 61 --IEBKGEENKPTFDVSKKKNPNQVNSQLNSHRSKDLQRE-----EHSOKS 107
Db 617 IVTKSESEKKNKAYDETNNKIKKSKNDLVLPEKKIKKEIKNEYIIDENKRGKQTEKGEKE 676
QY 108 DS-----TKDVTATVLDKNNI-----SSKSTTNNPK 134
Db 677 NKNNSIKKYKLTQDFEHLKEDNIEIGKKNESSTKKNNSNK 720

RESULT 36
Q8IC27_PLAF7
ID Q8IC27_PLAF7 PRELIMINARY; PRT; 1859 AA.
AC Q8IC27;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein PF07_0016.
GN Name=PF07_0016;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844506; CAD50814.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 1859 AA; 218375 MW; 2266544164BD360C CRC64;
Query Match 14.2%; Score 99; DB 2; Length 1859;
Best Local Similarity 24.7%; Pred. No. 1.2e+02;
Matches 36; Conservative 24; Mismatches 60; Indels 26; Gaps 5;
QY 6 TIVSEDFILPVYKGELEGYQFDGWEISGP-----EGKKDAGYVINSKDT- 52
Db 216 TDYSEDEIYAKYIQDKSSDSDSYQGYDKSLGKINTSNINMLNVTNKNQVNH--SMSNTI 273
QY 53 -----PIKVPFKIKIEKKEENKPTFDVSKKKNPNQVNSQLNSHRSKDLQREHSOKS 107
Db 274 QQDLSEFIHSINKYEKKKEKKNK-NYDRNKKSSNTNDKSYNTQDPRKNNQKEFVDNN 332
QY 108 DSTKDVATVLDKNNISSKSTTNNPN 133
Db 333 NKRND-----HNKNNELEQVYNNPN 353

RESULT 37
Q8I3A0_PLAF7
ID Q8I3A0_PLAF7 PRELIMINARY; PRT; 2563 AA.
AC Q8I3A0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE E1-E2 ATPase/hydrolase, putative.
GN Name=PF10240c;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Gobie A., Goodhead I., Gwilliam R., Horrocks P.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Johnson D., Lennard N.,
RA Humphray S., Jagels K., James K.D., Jarke N., Lawson D., Lawson N.,
RA Knights A., Konfortov B., Kyes S., Larke N., Mooney P., Moule S., Murphy L.,
RA Line A., Maddison M., McLean J., Mooney P., Moulé M.A., Rabbittowitsch E.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rajandream M.A.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Thayer A., Unwin L., Whitehead S., Woodward J.,
RA Suleston J.E., Craig A., Newbold C., Barrell B.G.;
RA "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002)
DR EMBL; AL829355; CAD51734.1; -; Genomic_DNA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:000524; F:ATP binding; IEA.
DR GO; GO:0015662; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0016820; F:hydrolase activity, acting on acid anhydrid. . .; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0008152; F:metabolism; IEA.
DR GO; GO:0030001; P:metal ion transport; IEA.
DR InterPro; IPR001757; ATPase E1-E2.
DR InterPro; IPR005834; Dehal like hydro.
DR InterPro; IPR008250; E1-E2_ATPase_reg.
DR InterPro; IPR008121; HeavyWe_transpt.
DR Pfam; PF00122; E1-E2_ATPase; 1.
DR Pfam; PF00702; Hydrolase; 2.
DR PRINTS; PR00119; CATATPASE.
DR TIGRFAMs; TIGR01494; ATPase P-type; 1.
DR PROSITE; PS00154; ATPASE_E1_E2; UNKNOWN_1.
DR PROSITE; PS50846; HMA_2; 1.
KW Hydrolase.
SQ SEQUENCE 2563 AA; 298806 MW; 3F9613243D26F8F1 CRC64;
Query Match 14.2%; Score 99; DB 2; Length 2563;
Best Local Similarity 29.4%; Pred. No. 1.6e+02;
Matches 35; Conservative 18; Mismatches 60; Indels 6; Gaps 4;
QY 20 GELEKGYQFDG-WEISGFGKKDAGYVINSKDTFTKPVFKIKIEKKEENKPTFDVSKK 78
Db 856 GEEBK-G-NIDGIYILKQNHKKOMIKGEENKONFNSKKEEKSDNSENEDKNYNLKR 914
QY 79 KD----NQVNSQLNSHRSKDLQRE- HSOKSDSTKDVATVLDKNNISSKSTTNNPN 133
Db 915 KEKHNSNDNDINDSVLKNVKEEETHESSESSNEQSDSYLKKIEKKDKNNISVDNNEN 973

RESULT 38
Q7RKU2_PLAYO
ID Q7RKU2_PLAYO PRELIMINARY; PRT; 393 AA.
AC Q7RKU2;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)

```

DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Drosophila melanogaster CG12259 gene product.  
 OS Name=PY02808;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=17XNL;  
 RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;  
 RA Carlton J.M., Anguilo S.V., Suh B.B., Kooij T.W., Perteau M.,  
 RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,  
 RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;  
 RT "Genome sequence and comparative analysis of the model rodent malaria  
 parasite Plasmodium yoelii yoelii.";  
 RL Nature 419:512-519(2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 CC  
 CC  
 CC  
 DR EMBL; AABL01000781; EAA22302.1; -; Genomic\_DNA.  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR InterPro; IPR007005; XAP5.  
 DR PANTHER; PTHR12722; XAP5; 2.  
 DR Pfam; PF04921; XAP5; 1.  
 SQ SEQUENCE 393 AA; 46652 MW; 55B30519B8FA97D2 CRC64;  
  
 Query Match 14.2%; Score 98.5; DB 2; Length 393;  
 Best Local Similarity 32.7%; Pred. No. 24;  
 Matches 35; Conservative 14; Mismatches 47; Indels 11; Gaps 5;  
  
 QY 36 FEGKDGAGVNLKSDT-----FIKPVFKIEKKEENKPTPDVSKKDNPOVNHQS 88  
 DB 90 FSKEDITY-NESSKDTKHKKHTNFKLSPFSDEDEDEDEKDNKSETPK-NKSD 147  
 QY 89 LNESHRKEDLQREHSQKSDSTKVTVTLVDKNNTSSKS-TYNNPNK 134  
 DB 148 EN-SLEKEQNEKEAEKSNTEQINKNTYDKNLQNGKSVNTENKNK 193  
  
 RESULT 39  
 Q7RL87 PLAYO  
 ID Q7RL87\_PLAYO PRELIMINARY; PRT; 674 AA.  
 AC Q7RL87;  
 DT 01-MAR-2004 (TREMBlrel. 26, Created)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Nochless-related.  
 GN Name=PY02598;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=73239;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=17XNL;  
 RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;  
 RA Carlton J.M., Anguilo S.V., Suh B.B., Kooij T.W., Perteau M.,  
 RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,  
 RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;

RT "Genome sequence and comparative analysis of the model rodent malaria  
 parasite Plasmodium yoelii yoelii.";  
 RL Nature 419:512-519(2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 CC  
 CC  
 CC  
 DR EMBL; AABL01000712; EAA22065.1; -; Genomic\_DNA.  
 DR HSRP; P16649; 1ERJ.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 8.  
 DR PRINTS; PR00320; GPROTEINERPT.  
 DR PRODOM; PD000018; WD40; 4.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 2.  
 DR PROSITE; PS00882; WD\_REPEATS\_2; 6.  
 DR PROSITE; PS0294; WD\_REPEATS\_REGION; 2.  
 KW Repeat; WD repeat.  
 SQ SEQUENCE 674 AA; 76535 MW; E7521B469FB8E0P7 CRC64;  
  
 Query Match 14.2%; Score 98.5; DB 2; Length 674;  
 Best Local Similarity 24.4%; Pred. No. 43;  
 Matches 30; Conservative 22; Mismatches 44; Indels 27; Gaps 5;  
  
 QY 15 LPVYKGLKGYQFDGWEISGFEGKDGAGVNLKSDTFIK-----PVFKIEE--- 63  
 DB 524 LAVFRGHVGPAYKI-AWSI-----DNNYIISQSDSTLKLWRINHLVPLKKKBE 575  
 QY 64 -----KKKEENKPTPDV--SKKDNPOVNHSQLNESHRKEDLQREHSQKSDSTKV 115  
 DB 576 QPKSDPKNDQKNDQKDKDKDDQKNDKNDHANNQDEGEKKKKKEKKKKIKK 635  
 QY 116 TVL 118  
 DB 636 TLL 638  
  
 RESULT 40  
 Q55G46 DICDI  
 ID Q55G46 DICDI PRELIMINARY; PRT; 827 AA.  
 AC Q55G46;  
 DT 13-SEP-2005 (TREMBlrel. 31, Created)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=DD0189573;  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AX4;  
 RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,  
 RA Sugrang R., Berriman M., Song J., Olsen R., Szafranski K., Xu O.,  
 RA Tungal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,  
 RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,  
 RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,  
 RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,  
 RA Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,  
 RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,  
 RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,  
 RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,  
 RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,  
 RA Loussich H., Mungall K., Oliver K., Price C., Quail M.A.,  
 RA Urushihara H., Hernandez J., Rabinowitch E., Steffen D., Sanders M.,  
 RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,  
 RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,  
 RA Shalisky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,  
 RA Chisholm R.L., Gibbs R., Loomis W.F., Platzner M., Kay R.R.,  
 RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.,  
 RT "The genome of the social amoeba Dictyostelium discoideum.";  
 RL Nature 0:0-0(2005).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.

DR EMBL; AAFI01000009; EAL73367.1; -: Genomic\_DNA.

KW Hypothetical protein; Nucleotide-binding.

SQ SEQUENCE 827 AA; 92700 MW; D576462666E4469 CRC64;

Query Match 14.2%; Score 98.5; DB 2; Length 827;  
Best Local Similarity 26.9%; Pred. No. 53;  
Matches 29; Conservative 17; Mismatches 31; Indels 31; Gaps 4;

QY 57 VPKIIEKK--EEENKPTFDVSKKKNPQVNHSQLNE--SHRKEDLQREH----- 103

Db 359 VIKRLERKKQKEBNSNKFSGGKKKKSKYNSSESESDSDSDIDFDDHHNHKKSKI 418

QY 104 -----SQKSDSTKQVT-----ATVLDKNNISSKSTTNPN 133

Db 419 KNISKEKNLSSNEHSDNEKDVNRNKIIQTQLDNNNNNNNNNNNN 466

Search completed: April 24, 2006, 14:59:34

Job time : 83.3893 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:50:52 ; Search time 14.0895 Seconds  
(without alignments)  
915.083 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773  
Perfect score: 696  
Sequence: 1 KEMSSSTIVSEEDFILPVYKG.....ATVLDKNNISSKSTNNPNK 134

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_80:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	696	100.0	2140	2 F95074 serine proteinase,
2	693	99.6	2144	2 A97942 metalloproteinase,
3	110	15.8	558	2 T18467 hypothetical prote
4	108.5	15.6	1038	2 JC5497 claustrin - chicke
5	99	14.2	211	2 T25911 hypothetical prote
6	98	14.1	665	2 B71609 hypothetical prote
7	97.5	14.0	1345	2 S46817 hypothetical prote
8	95	13.6	348	2 I37271 cylicin II - human
9	95	13.6	622	2 A90570 lipoprotein (impor
10	94.5	13.6	312	2 G81339 probable membrane
11	94	13.5	210	2 T28771 hypothetical prote
12	94	13.5	535	2 T37189 hypothetical prote
13	93	13.4	1397	2 T10466 DNA topoisomerase
14	92.5	13.3	219	2 B72291 hypothetical prote
15	92.5	13.3	325	2 T18283 hypothetical prote
16	91	13.1	253	2 T32879 hypothetical prote
17	89.5	12.9	508	2 B81594 hypothetical prote
18	89.5	12.9	508	2 B86549 hypothetical prote
19	89.5	12.9	508	2 C72074 hypothetical prote
20	89	12.8	528	2 B96795 unknown protein F2
21	89	12.8	1888	2 T39009 hypothetical prote
22	89	12.8	3724	2 T18427 hypothetical prote
23	88.5	12.7	301	2 T33068 hypothetical prote
24	88.5	12.7	371	2 A71683 hypothetical prote
25	88.5	12.7	385	2 T20410 hypothetical prote
26	88.5	12.7	540	2 D86432 hypothetical prote
27	88.5	12.7	644	2 T47835 hypothetical prote
28	88.5	12.7	762	2 G88436 protein 104A8.13 [
29	88.5	12.7	791	2 T24435 hypothetical prote

30	88.5	12.7	2401	2 T28676 rhoptry protein -
31	88	12.6	276	2 T23451 hypothetical prote
32	88	12.6	700	2 S67610 probable membrane
33	87.5	12.6	1702	2 A41859 iGA-specific metal
34	87.5	12.6	2523	2 T18477 hypothetical prote
35	87	12.5	635	2 F71621 hypothetical prote
36	87	12.5	2464	1 Q8MSF1 microtubule-associ
37	86.5	12.4	670	2 T28391 ORF MSV230 hypothe
38	86.5	12.4	792	2 T49989 hypothetical prote
39	86.5	12.4	2485	1 H71621 serine/threonine-s
40	86	12.4	1016	2 T19006 ankyrin related pr
41	85	12.2	335	2 T33457 hypothetical prote
42	85	12.2	867	2 T27136 hypothetical prote
43	85	12.2	871	2 T27135 hypothetical prote
44	85	12.2	988	2 T14188 hypothetical prote
45	85	12.2	1166	2 H71609 hypothetical prote

ALIGNMENTS

RESULT 1

F95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: F95074  
R:Tetrelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
naon, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morriss  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95074  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2140 <KUR>  
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match 100.0%; Score 696; DB 2; Length 2140;

Best Local Similarity 100.0%; Pred. No. 4.3e-46;

Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGKDGAGVYINLSKDTFIKPVPKK 60

Db 1973 KEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGPEGKDGAGVYINLSKDTFIKPVPKK 2032

Qy 61 IEKKKEENKPTPDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120

Db 2033 IEKKKEENKPTPDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2092

Qy 121 NNISSKSTTNNPNK 134

Db 2093 NNISSKSTTNNPNK 2106

RESULT 2

A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bargett, S.; DeHoff, B.S.;  
e, R.; LeBlanc, D.J.; Lee, I.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: A97942  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2144 <KUR>  
A;Cross-references: UNIPROT:Q8DQP7; UNIPARC:UPI00000E3490; GB:AE007317; PIDN:AAK99365.1;  
C;Genetics:  
A;Gene: prtA  
C;Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 693; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 7.4e-46;  
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 60  
Db 1977 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 2036

Qy 61 IEKKKEENKPTFDVSKKQNPQNVHSQLNESHKEDLQREHHSQKSDSTKDVATVLDK 120  
Db 2037 IEKKKEENKPTFDVSKKQNPQNVHSQLNESHKEDLQREHHSQKSDSTKDVATVLDK 2096

Qy 121 NNISKSTTNNPNK 134  
Db 2097 NNISKSTTNNPNK 2110

RESULT 3  
T18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18467  
R;Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A;Reference number: Z18937  
A;Accession: T18467  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-558 <LAW>  
A;Cross-references: UNIPROT:077355; UNIPARC:UPI000017CC2B; EMBL:AL008970; NID:e1407852;  
C;Genetics:  
A;Map position: 3  
A;Introns: 84/1; 160/1  
A;Note: C0465C

Query Match 15.8%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 0.34;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

Qy 7 IVSEEDFILPVY-----KGELEKGYQFDGWEISGPEGKK---DAGYVINLSKDTFIKPV 57  
Db 60 ILGFEDDILYFCISQLKQSKK--KADGEDKYLNAKLLNLTGFIGNKKSDIFIEEL 117

Qy 58 PKKI--EKKKEE-----ENKPTPDVSK-KKDNQNVHSQLN-----SHRK 95  
Db 118 LELLNEEKKEEHIADTLNENK-TNDIKKVNENINENVNENKDISNKKDKEHVSHQN 176

Qy 96 E-----DLQREH-----SOKSDSTK-----DVTATVLDKNNISKSTTN 130  
Db 177 EHNINNVNLKKEKEYTDIQDKRKHRLSLSQKSDSYKKRPFNKRTSTIER-SLSNKRDE 235

Qy 131 NPNK 134  
Db 236 KTNK 239

RESULT 4  
JC5497  
claustrin - chicken  
N;Alternate names: keratan sulfate proteoglycan  
C;Species: Gallus gallus (chicken)  
C;Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 09-Jul-2004  
C;Accession: JC5497; PC4334; S37561

R;Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A;Title: Claustrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A;Reference number: JC5497; MUID:94157526; PMID:7906711  
A;Accession: JC5497  
A;Molecule type: mRNA  
A;Residues: 1-1038 <BURI>  
A;Cross-references: UNIPROT:Q90784; UNIPARC:UPI00000FC87F; EMBL:X67778; NID:G406318; PID:  
A;Accession: PC4334  
A;Molecule type: protein  
A;Residues: 79-83;299-412;485-502 <BUR2>  
A;Cross-references: UNIPARC:UPI000017BFF3; UNIPARC:UPI000017BFF4; UNIPARC:UPI000017BFF5  
A;Experimental source: brain  
C;Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervous  
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
P;267-270/Region: cell attachment (R-G-D) motif  
F;112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 15.6%; Score 108.5; DB 2; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 0.89;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

Qy 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFK- 59  
Db 595 KPETKTIVAEKDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKESVKKEVKA 646

Qy 60 KIEKKKEENKPTFDVSKKQNPQNVHSQLNESHKEDLQRE-----EHSQKSDSTKDV 113  
Db 647 KPKEKKDEKPKKKEVSKKEEKLJ---KKBEKPKKEDIKKEVKKEVKKKEKKEAKKEV 702

RESULT 5  
T25911  
hypothetical protein T23B3.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T25911  
R;Maggi, L.; Le, T.  
submitted to the EMBL Data Library, February 1997  
A;Reference number: Z20109  
A;Accession: T25911  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-211 <MAG>  
A;Cross-references: UNIPROT:P91498; UNIPARC:UPI000007D762; EMBL:U88309; PIDN:AAB42334.1;  
A;Experimental source: strain Bristol N2; clone T23B3  
C;Genetics:  
A;Gene: CESP:T23B3.5  
A;Map position: 1  
A;Introns: 30/2; 200/3

Query Match 14.2%; Score 99; DB 2; Length 211;  
Best Local Similarity 29.4%; Pred. No. 0.83;  
Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

Qy 37 EGKDKAGYVINLSKDTFIKPVFKIIEKKKEENKPTFDVSKKQNPQNVH-----SQLNES 92  
Db 69 EGEKKDGEKKSEKDKGDKGEEKDEKKDDEKKDDEKKDDEKKDDEKKDDEKKDDEKKD 128

Qy 93 HRKEDLQREHHSQKSDSTKDVATVLDKNNISKSTTNNPNK 134  
Db 129 EKDDKKDEKKDEKKDEKKDEKKDEKKDEKKDEKKDEKKDEKKDEKKDEKKDEKK 170

RESULT 6  
B71609  
hypothetical protein PF0680w - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C;Accession: B71609  
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;

C:GeneticB:

Query Match 13.6%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 3;

[illegible]

Qy	124	SSK	126
Db	314	DSK	316

RESULT 9  
A90570  
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
C/Species: Mycoplasma pulmonis  
C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
C/Accession: A90570  
R/Chambaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galissou, P.  
Nucleic Acids Res. 29, 2145-2153, 2001  
A/Title: The complete genome sequence of the murine respiratory pathogen My  
A/Reference number: A90512: MUID:21267165; PMID:115353084

A;/reclues: 1-622 <RUK>  
A;/Cross-references: UNIPROT:Q98QAL; UNIPARC:UPI00000C80A1; GB:AL445566; PID:g14089879;  
A;/Experimental source: strain UAB CT1P  
C;/Genetics:  
A;/Gene: MYPU\_4650  
A;/Genetic code: SGC3

Query Match 13.6%; Score 95; DB 2; Length 622;  
Best Local Similarity 25.4%; Pred. No. 5.6;  
Matches 34; Conservative 26; Mismatches 50; Indels 24; Gaps 5;

<b>Qy</b>	8	VSEDPILPVYKGELEKGYQFDGWEISGFEGKKDAGVINLSKDTPIKPVPKIEKKEE	67
	:	:::	:
	:	:	:
<b>Dd</b>	85	LSEDNIIPLSLNTIENNAKQLDRF-----VSKDKFKIKPQEINFSQTE	128

Qy	68 ENKPTPDVSKKKD---NPNVAHSQLNES--HRKEDLQREHSSQKSDSTKDVTATVLDKN 121 : : ::     : : : : : : : : : : : : : : : :
Db	129 Q-KITDISSKEDEKNRPNKDNENNNNSDDQNDELQRKNSDKLNDNVODEKANKENS 187 :

Qy	122 -NISSKSTTNNPNK	134
	:	
Dd	188 SNDSKEKDENTNK	201

RESULT 10  
G81339  
probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)  
C/Species: Campylobacter jejuni  
C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C/Accession: G81339  
R/Parikhall, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilli  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barre  
Nature 403, 655-668, 2000

C/Spectres: Camryloaacker jejunu  
C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C/Accession: G81339  
R;/Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead  
Nature 403, 665-668, 2000









Query Match 12.8%; Score 89; DB 2; Length 3724;  
Best Local Similarity 22.3%; Pred. No. 1.2e+02;  
Matches 29; Conservative 27; Mismatches 42; Indels 32; Gaps 5;  
  
QY 6 TIVSEDFILPVYKGLKGYQFDGWEISGFGKXKQAGVYVNSKDTFIKPVFKIERKK 65  
DB 1060 TWICDNNIEQNEENSKGVISGTDN---ENKND-----MENKN 1097  
QY 66 EENKPTFDVSKKKNQVNH-SQLNESHKEDLQREHSHKSDSKDTKVATVLDKNNIS 124  
DB 1098 DMEKKN--DMEKKNDIEKKNDMEKKNDMEKKNDMEKKNDME-----NENNWE 1148  
QY 125 SEKSTNNPNK 134  
DB 1149 NKSDIENENK 1158

## RESULT 23

T33068  
hypothetical protein C35E7.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T33068  
R:Graves, T.; McDonald, R.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of C. elegans cosmid C35E7.  
A:Reference number: Z21278  
A:Accession: T33068  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-301 <GRA>  
A:Cross-references: UNIPROT:O61765; UNIPARC:UPI0000074979; EMBL:AF067216; PIDN:AAC17524.  
A:Experimental source: strain Bristol N2; clone C35E7  
C:Genetics:  
A:Gene: CESP:C35E7.9  
A:Map position: 1  
A:Introns: 30/3; 193/1; 236/2

Query Match 12.7%; Score 88.5; DB 2; Length 301;  
Best Local Similarity 24.8%; Pred. No. 8.1;  
Matches 27; Conservative 17; Mismatches 38; Indels 27; Gaps 3;  
  
QY 25 GYQFDGWEISGEGKK-----DAGVYVNSKDTFIKPVFKI-----EB 63  
DB 10 GTTAGWILAGCGKXKXGKSTASAAAPKADSKPKPVENVKSKKSEKPEKKEBP 69  
QY 64 KKEENKPTFDVSKKKNQVNHSQLNESHKEDLQREHSHKSDSKDTKD 112  
DB 70 KKEBKKEKSKKSEKKDDKK-----EBAKKEDDKDEKKEDKEDKDD 112

## RESULT 24

A71683  
hypothetical protein RP278 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004  
C:Accession: A71683  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark, U.  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: A71683  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-371 <AND>  
A:Cross-references: UNIPROT:Q9ZDP9; UNIPARC:UPI00000C1161; GB:AJ235271; GB:AJ235269; NID  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: RP278

Query Match 12.7%; Score 88.5; DB 2; Length 371;  
Best Local Similarity 23.1%; Pred. No. 10;

Matches 34; Conservative 33; Mismatches 37; Indels 43; Gaps 9;  
  
QY 1 KEMSTIVSEDFILPVYKGLKGYQFDGWEISGEGKK-DAGVYVNSKDTFIKPVFK 59  
DB 192 EQLNKTILNEE-----FRKNLEQ-----EKKKIEGKKTQASEALNKK---LKIYK 234  
QY 60 KIEKKER-----ENKPTFDVSKKKNQVNH---HSQLNESHKEDLQREHSHKSDSK 111  
DB 235 QMDEREELFKLAELIPQY-----AQAINDHAKLYAKQYQTKIENDPNYKELEK 286  
QY 112 DVTATVLDKNNISSKSTT-----NNPN 133  
DB 287 E-----IVSKIEYKSKSTNDIIINPN 309

## RESULT 25

T20410  
hypothetical protein E02A10.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T20410  
R:Thomas, K.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19271  
A:Accession: T20410  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-385 <WIL>  
A:Cross-references: UNIPROT:Q93424; UNIPARC:UPI00000835C8; EMBL:Z81053; PIDN:CAB02877.1.  
A:Experimental source: clone E02A10  
C:Genetics:  
A:Gene: CESP:E02A10.2  
A:Map position: 5  
A:Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 12.7%; Score 88.5; DB 2; Length 385;  
Best Local Similarity 35.5%; Pred. No. 11;  
Matches 22; Conservative 14; Mismatches 19; Indels 7; Gaps 2;

QY 59 KKEBKKEENKPTFDVSKKKNQVNHSQLNESHKEDLQREHSHKSDSKDTKVATV 118  
DB 317 KKEBKKEEKEQKE--EVEKKEE-----EKKDEBPKEEKEEKEEKEEVEKSEK 369  
QY 119 DK 120  
DB 370 EK 371

## RESULT 26

DB6432  
hypothetical protein T518.14 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: DB6432  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso-  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
Nasser, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;  
Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: DB6432  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-540 <STO>  
A:Cross-references: UNIPROT:Q9SA84; UNIPARC:UPI00000ABF35; GB:AE005172; NID:94587525; E

C:Genetics:  
A:Map position: 1

```

Query Match          12.7%; Score 88.5; DB 2; Length 540;
Best Local Similarity 24.4%; Pred. No. 15;
Matches 31; Conservative 25; Mismatches 54; Indels 17; Gaps 4;

QY      2  EMSTVISEEDFILPVYKGLKGYQDQWEISGFEGKQAGYVNLKDTFIKVPFKK- 60
      | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      49  EKASPFKEEDFPFADLKESEK-----ALSLKSLKEAIVDN---TLTKKKES 96

QY      61  --IEKKKEENKPTFDVSKKKNQVNHSQLNESHRKEDLQREHSQKSDSTKDVATVL 118
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      97  SPNKEKKEEVVYKPAEVEKKKE--EAAAEKVEEKKSEAVVTEEAPKAEVTVTEII 154

QY      119 DKNNISS 125
      | : : :
Db      155 PKEEVT 161

RESULT 27
T47835
hypothetical protein T209.90 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C;Accession: T47835
R;Nykatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;
submitted to the Protein Sequence Database, February 2000
A;Reference number: Z24475
A;Accession: T47835
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-644 <NYA>
A;Cross-references: UNIPROT:Q9M1D2; UNIPARC:UPI0000090AD; EMBL:AL1138658
A;Experimental source: cultivar Columbia; BAC clone T209
C;Genetics:
A;Map position: 3
A;Introns: 158/2; 329/3
A;Note: T209.90

```

Query Match	12.7%; Score 88.5; DB 2; Length 644;
Best Local Similarity	24.6%; Pred. No. 19;
Matches	32; Conservative 18; Mismatches 53; Indels 27; Gaps 3;
Qy	9 SEBDFLLPVVKGELEKGVQPDGWEISGFEGKKDAGVYINLSKDTPIKPVPKIEBKKEE 68     : :     : :     : :     :
Dd	534 SVADFLLRIKIKNSPQK-----ETTSKNQKNDGNV-----KKEND 570     : :     : :     : :     :
Qy	69 NKPTFDVYSKKDNPNVNHSQLNESHRKEDLQREHSOKSDS----TKDVTATVLDDQNIS 124 : :     : :     : :     : :     :
Dd	571 HQKSGDNVKKENSVKVPRELRSTGGKKVEVENNSSSKRKQTETA EVAVGKGRE 630 : :     : :     : :     : :     :
Qy	125 SKSTTTNPNK 134 
Dd	631 SGKDDKQPRK 640 

RESULT 28  
G88436  
protein T04A8.13 [imported] - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C/Accession: G88436  
R;anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biological processes  
A;Reference number: A75000; PMID:9069613; PMID:9851916  
A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 283, 2103, 1999.  
A;Accession: G88436  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-762 <STO>  
A;Cross-references: UNIPROT:Q22142; UNIPARC:UPI0000017A5C4; GB:chr\_III; PIDN:CAA84732.1;  
C;Genetics:  
A;Gene: T04A8.13

[illegible]

RESULT 29  
T24435  
hypothetical protein T04A8.13 - *Caenorhabditis elegans*  
C/Species: *Caenorhabditis elegans*  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T24435  
R/Palmer, S.  
submitted to the EMBL Data Library, August 1994  
A/Reference number: Z19889  
A/Accession: T24435  
A/Status: preliminary; translated from JEB/EMBL/DDBJ  
A/Molecule type: DNA  
A/Residues: 1-791 <WIL>  
A/Cross-references: UNIPROT:Q22142; UNIPARC:UPI0000061132; EMBL:Z35663; PIDN:ZC  
A/Experimental source: clone T04A8  
C/Genetics:  
A/Gene: CESP:T04A8.13  
A/Map position: 3  
A/Introns: 31/3; 212/1; 229/3; 331/3; 406/1; 472/3; 572/1; 651/2

Query Match	12.7%;	Score 88.5;	DB 2;	Length 791;
Best Local Similarity	25.2%;	Pred. No. 24;		
Matches	32;	Conservative 25;	Mismatches 65;	Indels 5; Gaps 3
Qy	9	SEEDFILPVYKGELEKGYQFDGWEISGFEGK-KDAGYVINLSKDTFTKVPFKTKEEKKEE	67	
Db	14	SGQOIMTPGY--DKKEGLGMDQKEIVGDDKKDKKRRKRKLQDFAE--LKQDEEKDKKE	69	
Qy	68	ENKPTFDVVSQKONPQVNHSQLSHESHRKEDLQREHSQKSDSTKDVATVTLVDKNNISSKS	127	
Db	70	EAEKEKNEKEGEEKEEKDGHEKKEDKDKCKENENDEBKEKSCKDKKEESKEDCKCKTK	129	
Qy	128	TTNNPNK	134	
Db	130	TEDNEGK	136	

```

RESULT 30
T28676
  roptxy protein - Plasmodium yoelii (fragment)
C:Species: plasmodium yoelii
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676; A45521
E:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507; MUID:97077455; PMID:8920022
A:Accession: T28676
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: UNIPROT:Q6216; UNIPARC:UPI000017B647; EMBL:U36927; NID:g1041784; P:
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990

```

A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple cd  
A:Reference number: A45521; MUID:91101660; PMID:2270106  
A:Accession: A45521  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 2260-2401 <KER>  
A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

Query Match 12.7%; Score 88.5; DB 2; Length 2401;  
Best Local Similarity 28.4%; Pred. No. 81;  
Matches 38; Conservative 24; Mismatches 47; Indels 25; Gaps 5;

QY 9 SEEDTILPVYKKELEKGY--QPDGWEISGFGKKDAGVINLSKDTIKPVFKIKBEKK 65  
DB 947 SEESYINDI-KKELEKNVLESQNNNTDINQYLSKIENTY--NILKLNKIKKIIDKVKRYT 1003

QY 66 BRENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKD----- 112  
DB 1004 DIEK-----NNKNNAELNSSEKIITQLKENSLSKECOSKIKSTIIDNYVSECINKITN 1058

QY 113 -VTATVLDKNNISS 125  
DB 1059 LKTYIVNEKNINT 1072

RESULT 31  
T23451  
hypothetical protein K08E3.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T23451  
R:McMurray, A.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19743  
A:Accession: T23451  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-276 <WIL>  
A:Cross-references: UNIPROT:Q9XUT1; UNIPARC:UPI0000077D88; EMBL:Z81568; PIDN:CAB04590.1;  
A:Experimental source: clone K08E3  
C:Genetics:  
A:Gene: CESP:K08E3.2  
A:Map position: 3  
A:Introns: 81/3; 102/3; 169/3; 211/2

Query Match 12.6%; Score 88; DB 2; Length 276;  
Best Local Similarity 31.5%; Pred. No. 8;  
Matches 28; Conservative 10; Mismatches 29; Indels 22; Gaps 3;

QY 48 LSKDTFIKPVFKIKBEKKEE--NKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQ 105  
DB 60 LEKDEDVEKKKEGKEEKEEKKEDKBEKKD-----SKEDKKDEDHEK 107

QY 106 KSDSTKDVATVLDKNNISSKSTTNPNK 134  
DB 108 KK-----TAEKENNEKKDNKNK 128

RESULT 32  
S67610  
probable membrane protein YD1074c - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D2483  
C:Species: Saccharomyces cerevisiae  
C>Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: S67610  
R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67608  
A:Accession: S67610  
A:Molecule type: DNA  
A:Residues: 1-700 <WAM>  
A:Cross-references: UNIPROT:Q07457; UNIPARC:UPI0000069BFF; EMBL:Z74122; NID:g1431087; PI  
A:Experimental source: strain S288C

C:Genetics:  
A:Gene: SGD:BRE1; MIPS:YDL074c  
A:Cross-references: SGD:S0002232  
A:Map position: 4L  
C:Keywords: transmembrane protein  
P:69-85/Domain: transmembrane #status predicted <TMM>

Query Match 12.6%; Score 88; DB 2; Length 700;  
Best Local Similarity 23.7%; Pred. No. 22;  
Matches 31; Conservative 24; Mismatches 40; Indels 36; Gaps 5;

QY 21 ELEKGYQDGEWISGFGKK-----DAGVINL-SKDTF---IK 55  
DB 447 DLEKGFR---ELSDLTHKKYSRIINSHSVISKLTVEKTKADQKYFAAMRSKDSLIEIK 502

QY 56 PVFKKIEKKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTA 115  
DB 503 TLSKLSKSNEL-----ILQLKSDRLLOQKIGNLHKQLDLSQNNERRLLDSSKTETL 555

QY 116 TVLDKNNISSK 126  
DB 556 KIIDLNTSTK 566

RESULT 33  
A41859  
IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae  
C:Species: Haemophilus influenzae  
A:Variety: strain HK715  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: A41859  
R:Poulsen, K.; Reinholdt, J.; Kilian, M.  
J. Bacteriol. 174, 2913-2921, 1992  
A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae  
A:Reference number: A41859; MUID:92234949; PMID:1373717  
A:Accession: A41859  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1702 <POU>  
A:Cross-references: UNIPROT:P45384; UNIPARC:UPI000012D3F0; GB:M87489; NID:g148906; PIDN  
A:Experimental source: strain HK715  
A>Note: sequence extracted from NCHI backbone (NCBIP:97282)  
C:Superfamily: IGA-specific metalloendopeptidase  
C:Keywords: hydrolase; metalloproteinase

Query Match 12.6%; Score 87.5; DB 2; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 66;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 46 INLSKDTFIKPVFKIKBEKKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHS- 104  
DB 1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSSSESPKSRRRSI 1355

QY 105 --QKSDSTKDVATVLDKNNISSKSTTNPNK 134  
DB 1356 SQPQTSABETTAASTDDETTIADNSKRSKPNR 1387

RESULT 34  
T18477  
hypothetical protein C0485w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18477  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937  
A:Accession: T18477  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-2523 <LAW>  
A:Cross-references: UNIPROT:O77365; UNIPARC:UPI000017CC31; EMBL:AL008970; NID:e1407852;  
C:Genetics:





**THIS PAGE BLANK (0070)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 10.4922 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTIVSEEDFILPVYKG.....ATVLDKNINSSKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 225428 seqs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA\_New:\*  
1: /SID95/prodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /SID95/prodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /SID95/prodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /SID95/prodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
5: /SID95/prodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
6: /SID95/prodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /SID95/prodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /SID95/prodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	107	15.4	354	7	US-11-189-817-2
2	103.5	14.9	651	7	US-11-128-660-1
3	88	12.6	746	6	US-10-793-626-1780
4	86	12.4	443	7	US-11-188-298-1015
5	85	12.2	472	6	US-10-793-626-658
6	85	12.2	720	6	US-10-793-626-2058
7	84.5	12.1	140	7	US-11-096-568A-4771
8	84.5	12.1	501	6	US-10-485-517-381
9	84.5	12.1	645	6	US-10-485-517-244
10	84.5	12.1	1694	7	US-11-052-554A-83
11	82	11.8	439	7	US-11-188-298-15964
12	82	11.8	439	7	US-11-188-298-16606
13	82	11.8	886	7	US-11-087-099-11456
14	81.5	11.7	8746	7	US-11-098-686-10232
15	79.5	11.4	493	7	US-11-096-568A-3070
16	79.5	11.4	493	7	US-11-096-568A-3071
17	79.5	11.4	510	7	US-11-096-568A-3069
18	79.5	11.4	943	6	US-10-475-204-34
19	79.5	11.4	1036	7	US-11-096-568A-28315
20	79.5	11.4	1070	7	US-11-096-568A-28314
21	79.5	11.4	1276	7	US-11-096-568A-28313
22	79	11.4	732	7	US-11-124-368A-248
23	79	11.4	803	7	US-11-124-368A-241
24	79	11.4	803	7	US-11-124-368A-242
25	78.5	11.3	258	7	US-11-098-686-10475

26	78.5	11.3	700	7	US-11-196-475-74	Sequence 74, Appl
27	78.5	11.3	1155	6	US-10-793-626-1780	Sequence 1780, Ap
28	78	11.2	191	6	US-10-860-601-5	Sequence 5, Appl
29	77.5	11.1	405	6	US-10-485-517-239	Sequence 239, App
30	77.5	11.1	627	6	US-10-873-528-191	Sequence 191, App
31	77	11.1	568	6	US-10-793-626-2482	Sequence 2482, Ap
32	77	11.1	663	7	US-11-196-475-70	Sequence 70, Appl
33	77	11.1	663	7	US-11-196-475-78	Sequence 78, Appl
34	77	11.1	1145	6	US-10-793-626-1432	Sequence 1432, Ap
35	76.5	11.0	439	7	US-11-188-298-17915	Sequence 17915, A
36	76	10.9	244	7	US-11-096-568A-11931	Sequence 11931, A
37	76	10.9	482	7	US-11-232-440-33	Sequence 33, Appl
38	76	10.9	502	7	US-11-045-004-1651	Sequence 1651, Ap
39	75.5	10.8	278	6	US-10-793-626-2700	Sequence 2700, Ap
40	75.5	10.8	404	6	US-10-793-626-398	Sequence 398, App
41	75.5	10.8	558	6	US-10-510-903-22	Sequence 22, Appl
42	75	10.8	513	7	US-11-185-924-16	Sequence 16, Appl
43	74.5	10.7	313	7	US-11-096-568A-32043	Sequence 32043, A
44	74.5	10.7	425	7	US-11-096-568A-32042	Sequence 32042, A
45	74.5	10.7	434	7	US-11-096-568A-32041	Sequence 32041, A

## ALIGNMENTS

RESULT 1  
US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US20060030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRUIIHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM  
; TITLE OF INVENTION: MSP3 TARGETED BY BIOLOGICALLY ACTIVE ANTIBODIES  
; FILE REFERENCE: 275601US0  
; CURRENT APPLICATION NUMBER: US/11/189,817  
; PRIOR FILING DATE: 2005-07-27  
; PRIOR APPLICATION NUMBER: 60/598,062  
; PRIOR FILING DATE: 2004-08-03  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 2  
; LENGTH: 354  
; TYPE: PRT  
; ORGANISM: Plasmodium falciparum  
US-11-189-817-2

Query Match 15.4%; Score 107; DB 7; Length 354;  
Best Local Similarity 23.8%; Pred. No. 0.047;

Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;

Qy 1 KEMSTIVSEEDFILPVYKGEGYQPDGWEISGP--EGKCDAG-----TVINLSKDTF 53

Db 185 KEASS-----VDYIL-----GWFQGGVPEHKKEKRMMLSHLYVSSKDKNI 225

Qy 54 IKPVFKIEEKKEB-----ENKPTFDVSKKONQVNHQSOLNESHKEDIQR 100

Db 226 SKENDVDLDEKEAEETEEELKEKKEETESISEDEEEEEEKEEENDKKQEK 285

Qy 101 BEHSQKSDSTKDVATVLDKNINSSKSTNN 131

Db 286 EQSNENNDQKKDMEA-----QNLISKQNQNN 311

## RESULT 2

US-11-128-660-1

; Sequence 1, Application US/11128660

; Publication No. US20060024324A1

; GENERAL INFORMATION:

; APPLICANT: Statens Serum Institut

; TITLE OF INVENTION: Vaccines comprising chimeric malaria proteins derived from Plasmodium falciparum

; FILE REFERENCE: 15007dk

```
; CURRENT APPLICATION NUMBER: US/11/128,660
; CURRENT FILING DATE: 2005-05-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 651
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-11-128-660-1

Query Match      14.9%; Score 103.5; DB 7; Length 651;
Best Local Similarity 25.2%; Pred. No. 0.19;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQDGEISGF--EGKKDAG-----YVINLSKDTF 53
DB 483 KEASS-----YDIIL-----GWFFGGVPEHKKEENMLSHLYVSSKKKENI 523

QY 54 IKPVFKKIEKKERENKPTFVSKKKNPQVNHSQLN-----ESHKKEDLQR 100
DB 524 SKENDVDLDB--KEEAEATEBELEKKEETESEISEDEEEEEEEKEEENEKKGEQEK 582

QY 101 EHSQKSDSTKDVATVLDKNNISSKSTNN 131
DB 583 EQSNENNQKKDMEA-----QNLISKQNNN 608

RESULT 3
US-10-793-626-652
; Sequence 652, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR FILING DATE: 2004-03-04
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 652
; LENGTH: 746
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-652

Query Match      12.6%; Score 88; DB 6; Length 746;
Best Local Similarity 25.7%; Pred. No. 5;
Matches 37; Conservative 21; Mismatches 50; Indels 36; Gaps 7;

QY 4 SSTIVSEEDFILPVYKGELEKGYQDGW---EISGF-----GKKDAGYVIN--LSKD 51
DB 611 SGTKVLPHSKVMLMDGELTMP-DMTGWTKEVDLAFELDTKLKVTGKNGFVTQISKG 669

QY 52 TPIKPVFKKIEKKERENKPTFVGS-----KKKNPQVNHSQLNESHRKEDLQREHSQKS 107
DB 670 QIHK-----NKKIEVLSAEDTDDQKTDDESDDNKSKKDKADEHSNTS 716

QY 108 DSTKDVATVLDKNNISSKSTNN 131
DB 717 SSTKN-----DKSNADSKNDSDD 734

RESULT 4
US-11-188-298-1015
; Sequence 1015, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.

; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 1015
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Thermococcus kodakaraensis
US-11-188-298-1015

Query Match      12.4%; Score 86; DB 7; Length 443;
Best Local Similarity 17.8%; Pred. No. 4.1;
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

QY 18 YKGELEKGYQDGEISGFEGKKDAGYVINLSKDTFI----- 54
DB 42 YEEAVEGVSFDGSSIPGFGIEDSLIPKADPSTYAEIPWEGIGRVGVYIKGDEPYQA 101
QY 55 --KEVFKKIEKKEREN-----KPTFDVSKKKD----- 80
DB 102 DPGILKRVLERLEKEGLKAHIGPEFYIFKNGTWTWELHIPDSGGYFDLVGLDKAREIR 161
QY 81 -----NPOVNHSQLNESHRKEDLQREHSQKSD---STKDVATVLD 119
DB 162 REIALYMPYLGKPEVLHHEVGKAQHEIDFRYDEALRADNIVSFKHVVKAAVE 215

RESULT 5
US-10-793-626-658
; Sequence 658, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-658

Query Match      12.2%; Score 85; DB 6; Length 472;
Best Local Similarity 32.6%; Pred. No. 5.4;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 61 IEKKERENKPTFVSKKKNPQVNHSQLNESHRKEDLQ-----REEHSQKSDS--- 109
DB 3 MEENKNQPNKE--NMSKNKDNA-----THLADSHREDELELFRRNKNARQRRRRIDNQSK 56

QY 110 TKDVAT-----VLDKNNISSKSTNNPK 134
DB 57 EKDATSTQSQLETKPMDFLDNHNKS--HNQNK 86

RESULT 6
US-10-793-626-2058
; Sequence 2058, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
```

```
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793.626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2058

Query Match          12.1%; Score 85; DB 6; Length 720;
Best Local Similarity 28.0%; Pred. No. 8.8;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

Qy 61 IEEKKEBENKPTPDVSKKD----NPQVNHSQLNESHKEDLQREHHSQKSDSTK----D 112
Db 38 LEEQIKALDKFKKASQAKDTNKQNTQNNHQSNNKQNSNDKEKQSQKNSKPTKKKEQN 97

Qy 113 VTATVLDKNNISSKSTTNNPK 134
Db 98 NKGQKQNNKNTNKQNNK 119

RESULT 7
US-11-096-568A-4771
; Sequence 4771, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096.568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4771
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(140)
; OTHER INFORMATION: Ceres Seq. ID no. 14304111
US-11-096-568A-4771

Query Match          12.1%; Score 84.5; DB 7; Length 140;
Best Local Similarity 26.6%; Pred. No. 1.4;
Matches 33; Conservative 25; Mismatches 39; Indels 27; Gaps 6;

Qy 10 BEDFILPVY--KGELEKGYQDQWEGISGFGKQDAGYVYNLSKDTFKVPFKIEKKKE 67
Db 23 EDEEKIPAYRRGRGPQPMKD-----FEEDEE-----BELVKMEKEEEE 65

Qy 68 ENKPTFDVSKKQNPQ-----VNHSQLNESHKEDLQREHHSQKSDSTKQVATVLDKNNI 123
Db 66 DDSVT---SKKEBERKRMKMGNSDANKEKGLGSKSRDGSSTKS-TSTGFRQNGS 121

Qy 124 SSKS 127
Db 122 RRKS 125

RESULT 8
US-10-485-517-381
; Sequence 381, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
```

```
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485.517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-381

Query Match          12.1%; Score 84.5; DB 6; Length 501;
Best Local Similarity 28.6%; Pred. No. 6.4;
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;

Qy 41 DAGYVINL-SKDTFIKPVFKIEKKEBENKPTFDV-----SKKQNPQVNHSQLNESH 94
Db 297 DGQYHVRIVDKAEFTKANTDKSNKKEQDNSAKKEATPATPSKPTSPFVKESQKQDSQK 356

Qy 95 KEDLQ----REHHSQKSDSTKQV--ATVLDKNNISSKSTTNNPK 134
Db 357 DDNKQLPSVEKENDASSESGDKTTPATKPTKGEVSSSTT--PTK 399

RESULT 9
US-10-485-517-244
; Sequence 244, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485.517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 244
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-244

Query Match          12.1%; Score 84.5; DB 6; Length 645;
Best Local Similarity 28.6%; Pred. No. 8.6;
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;

Qy 41 DAGYVINL-SKDTFIKPVFKIEKKEBENKPTFDV-----SKKQNPQVNHSQLNESH 94
Db 441 DGQYHVRIVDKAEFTKANTDKSNKKEQDNSAKKEATPATPSKPTSPFVKESQKQDSQK 500

Qy 95 KEDLQ----REHHSQKSDSTKQV--ATVLDKNNISSKSTTNNPK 134
Db 501 DDNKQLPSVEKENDASSESGDKTTPATKPTKGEVSSSTT--PTK 543

RESULT 10
US-11-052-554A-83
```

```
; Sequence 83, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 83
; LENGTH: 1694
; TYPE: PRT
; ORGANISM: Haemophilus influenzae Rd
US-11-052-554A-83

Query Match      12.1%; Score 84.5; DB 7; Length 1694;
Best Local Similarity 26.1%; Pred. No. 27;
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

Qy 46 INLSKDTFTKPVFKKIEEKKEENKPTFDVSKKKNQVNHSQLNESHKEDLQREHS- 104
Db 1288 INTGATATETAKSDKQFETAASTEDAQHKANTVADNSVANNSESSDPKSRRSI 1347

Qy 105 --QKSDTKDVTATVLDKNNISSKSTNNPNK 134
Db 1348 SQPQTSABETTAASTDETTIADNKRKSPNR 1379
```

```
RESULT 11
US-11-188-298-15964
; Sequence 15964, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; PRIOR FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 15964
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus DSM 3638
US-11-188-298-15964

Query Match      11.8%; Score 82; DB 7; Length 439;
Best Local Similarity 18.1%; Pred. No. 9;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

Qy 25 GYQFGWEISGPEGKKGAGYVNLKDTFI-----KP-----VFK 59
Db 45 GISFDGSSVPGFQIGIEDSLVFKADPTDYVEVPWPNVAVGYFYKKNKPYGADPRGILK 104

Qy 60 KIEKKKEEN-----KPTFDVSKKKN-----81
Db 105 RALEBEKEGYKAYIGPEPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIAYM 164

Qy 82 -----POVNHSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 123
Db 165 PSFGLIPEVLHHEVCKAQHEIDFRYDEALKTADNIVSFKYITKVAEMHGL 215
```

```
RESULT 12
US-11-188-298-16606
; Sequence 16606, Application US/11188298
```

```
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; PRIOR FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 16606
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus woesei
US-11-188-298-16606

Query Match      11.8%; Score 82; DB 7; Length 439;
Best Local Similarity 18.1%; Pred. No. 9;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

Qy 25 GYQFGWEISGPEGKKGAGYVNLKDTFI-----KP-----VFK 59
Db 45 GISFDGSSVPGFQIGIEDSLVFKADPTDYVEVPWPNVAVGYFYKKNKPYGADPRGILK 104

Qy 60 KIEKKKEEN-----KPTFDVSKKKN-----81
Db 105 RALEBEKEGYKAYIGPEPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIAYM 164

Qy 82 -----POVNHSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 123
Db 165 PSFGLIPEVLHHEVCKAQHEIDFRYDEALKTADNIVSFKYITKVAEMHGL 215
```

```
RESULT 13
US-11-087-099-11456
; Sequence 11456, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 11456
; LENGTH: 886
; TYPE: PRT
; ORGANISM: Bacillus cereus ATCC 14579
US-11-087-099-11456

Query Match      11.8%; Score 82; DB 7; Length 886;
Best Local Similarity 24.4%; Pred. No. 20;
Matches 40; Conservative 25; Mismatches 53; Indels 46; Gaps 8;

Qy 11 EDFTLIPVY--KGEL-----EKGYQF--DGWEIS-GFEGKGD 41
Db 557 EEIQIPVYDLEGESEIENQLVSEGGTFNGVKGWSTFGEKYVYKFDLSDEISIRNGI-- 614

Qy 42 AGYVINLSKDTFTKPVFKKIEEKKEENKPTFDVSKKKNQVNHSQLNESHKEDLQ-- 99
Db 615 --VIQNIVEKEEKEPEVKIEETKEEKEPEVKIEETKEEVPKVEET--KEEVKEP 670

Qy 100 ----REEHSQKSDTKDVTATVLD-----KNNISSKSTNNPNK 134
Db 671 TKVEEAKKEVKPTKEVETKEEVPKVEEAKKEVKEPTK 714
```

```
RESULT 14
US-11-098-686-10232
; Sequence 10232, Application US/11098686
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
```

```
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10232
; LENGTH: 8746
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10232

Query Match      11.7%; Score 81.5; DB 7; Length 8746;
Best Local Similarity 25.9%; Pred. No. 3.3e+02;
Matches 38; Conservative 24; Mismatches 60; Indels 25; Gaps 7;

Qy  2 EMSSTIVSEEDFIL--PVYKGELEKGYQFDGWE--ISGFEGKKGADGYYVNLKSDTFIKPV 57
Db  3358 EGEDVAEEEDILLWDLTLDNEAEGTKEEHAEVKVGEVGEVFDG-----ISEED--KP- 3410

Qy  58 FKIEEKEEENKPTFDVSKKKNPQVNHSQLNESHKEDL-----ORBEHSQKS 107
Db  3411 --KKDKEEQKATGLDSSGETTEESQPOQEBEKEKENSFGSNESPSPQOEBSYDE 3468

Qy  108 DSTKDVATVLDKNNISSKSTTNNPNK 134
Db  3469 TSSVVTSSPLLSINEV--KQTEDEKSAK 3493

RESULT 15
US-11-096-568A-3070
; Sequence 3070, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3070
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc feature
; LOCATION: (1)-(493)
; OTHER INFORMATION: Ceres Seq. ID no. 15172486
US-11-096-568A-3070

Query Match      11.4%; Score 79.5; DB 7; Length 493;
Best Local Similarity 25.5%; Pred. No. 17;
Matches 40; Conservative 19; Mismatches 59; Indels 39; Gaps 7;

Qy  4 SSTIVSEE-----DFILPVYKGELEKGYQFDGWEISGFEGKKGADGVI----- 46
Db  90 ANTIIVTEEKPGKVPVKKKIKTPVSK---KKDETADSNKTTETLSDDKDEGNVAVQAQDD 146

Qy  47 -----NLKSDTFIKVPFKIEB--KKEEENKPTFDVSKKKNPQVNHSQLNESHKED 97
Db  147 TQSTGKQTANADTTVTPEVKTKGVKVPKQSKTPT---SEKRDN--TADSSKTETKSDKD 201

Qy  98 LQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 134
Db  202 DKKEER-----VTGEKSGAKTDKLKASDKDVTNVKKG 233

RESULT 16
US-11-096-568A-3071
; Sequence 3071, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3071
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc feature
; LOCATION: (1)-(493)
; OTHER INFORMATION: Ceres Seq. ID no. 16625552
US-11-096-568A-3071

Query Match      11.4%; Score 79.5; DB 7; Length 493;
Best Local Similarity 25.5%; Pred. No. 17;
Matches 40; Conservative 19; Mismatches 59; Indels 39; Gaps 7;

Qy  4 SSTIVSEE-----DFILPVYKGELEKGYQFDGWEISGFEGKKGADGVI----- 46
Db  90 ANTIIVTEEKPGKVPVKKKIKTPVSK---KKDETADSNKTTETLSDDKDEGNVAVQAQDD 146

Qy  47 -----NLKSDTFIKVPFKIEB--KKEEENKPTFDVSKKKNPQVNHSQLNESHKED 97
Db  147 TQSTGKQTANADTTVTPEVKTKGVKVPKQSKTPT---SEKRDN--TADSSKTETKSDKD 201

Qy  98 LQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 134
Db  202 DKKEER-----VTGEKSGAKTDKLKASDKDVTNVKKG 233

RESULT 17
US-11-096-568A-3069
; Sequence 3069, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3069
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc feature
; LOCATION: (1)-(510)
; OTHER INFORMATION: Ceres Seq. ID no. 15172485
US-11-096-568A-3069

Query Match      11.4%; Score 79.5; DB 7; Length 510;
Best Local Similarity 25.5%; Pred. No. 18;
Matches 40; Conservative 19; Mismatches 59; Indels 39; Gaps 7;

Qy  4 SSTIVSEE-----DFILPVYKGELEKGYQFDGWEISGFEGKKGADGVI----- 46
Db  107 ANTIIVTEEKPGKVPVKKKIKTPVSK---KKDETADSNKTTETLSDDKDEGNVAVQAQDD 163

Qy  47 -----NLKSDTFIKVPFKIEB--KKEEENKPTFDVSKKKNPQVNHSQLNESHKED 97
Db  164 TQSTGKQTANADTTVTPEVKTKGVKVPKQSKTPT---SEKRDN--TADSSKTETKSDKD 218
```

```
Qy 98 LQREHSQKSDSTKDVDTATVLDKNNISSKSTNNPNK 134
      :||| : : : ||| : ||| : |||
Db 219 DKKEER-----VTGKSGAKTDKLKASDKDVTNVGK 250
      :||| : : : ||| : ||| : |||

RESULT 18
US-10-475-204-34
; Sequence 34, Application US/10475204
; Publication No. US20050277116A1
; GENERAL INFORMATION:
; APPLICANT: PRESIDENT AND FELLOWS OF HARVARD COLLEGE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE IDENTIFICATION OF
; FILE OF INVENTION: PROTEIN INTERACTIONS IN VERTEBRATE CELLS
; FILE REFERENCE: HMV-056.25
; CURRENT APPLICATION NUMBER: US/10/475,204
; CURRENT FILING DATE: 2003-10-17
; PRIOR APPLICATION NUMBER: PCT/US02/13008
; PRIOR FILING DATE: 2002-09-27
; PRIOR APPLICATION NUMBER: 60/285,509
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-475-204-34

Query Match 11.4%; Score 79.5; DB 6; Length 943;
Best Local Similarity 23.1%; Pred. No. 36;
Matches 36; Conservative 16; Mismatches 55; Indels 49; Gaps 5;

Qy 16 PVYKGELEKGYQFDGWEISGEGKKDAGVYVNLKDTFKVPFKIEKK-----EEN 69
      :||| : : : ||| : ||| : |||
Db 375 PSDKTLVTSYALIDETVNNYRSTKYEMYSKNAEPPSRSKRTIKQKRRKFMKPAERQ- 433
      :||| : : : ||| : ||| : |||

Qy 70 KPTFDVSKKCD-NPQVNHSQLNESHKEDLQREEH-----SOKSDSTK 111
      :||| : : : ||| : ||| : |||
Db 434 ---LDVGSKDENIHTSHITQDEFNRSDRNMEEHENGDCVKKQMPVGGKSKSTRK 490
      :||| : : : ||| : ||| : |||

Qy 112 D-----VTATVLDKNNISSK 126
      :||| : : : ||| : ||| : |||
Db 491 DKESKKKRFSSSEKKNLVPERTVTSTVKRSRISR 526
      :||| : : : ||| : ||| : |||

RESULT 19
US-11-096-568A-28315
; Sequence 28315, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28315
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1036)
; OTHER INFORMATION: Ceres Seq. ID no. 2712010
US-11-096-568A-28315

Query Match 11.4%; Score 79.5; DB 7; Length 1036;
Best Local Similarity 23.2%; Pred. No. 41;
Matches 23; Conservative 20; Mismatches 47; Indels 9; Gaps 2;

Qy 37 EGKKDAGVYVNLKDTF-----IKPVFKIEEKEEENKPTFDVSKKKNPQVNHSQLN 90
      :||| : : : ||| : ||| : |||
```

```
Db 552 DGDSDKKVMEVGKSSDSGSVEMKPTAESLIEDVKDENASKTVDVKQETGSPDTKKKEGA 611
      :||| : : : ||| : ||| : |||
Qy 91 ESHRKEDLQREEH---SQKSDSTKDVDTATVLDKNNISSK 126
      :||| : : : ||| : ||| : |||
Db 612 SSSSKDQTKTGEDKKAEEKNNSETMTMSEGGKIDRNNITDEK 650
      :||| : : : ||| : ||| : |||

RESULT 20
US-11-096-568A-28314
; Sequence 28314, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28314
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1070)
; OTHER INFORMATION: Ceres Seq. ID no. 2712009
US-11-096-568A-28314

Query Match 11.4%; Score 79.5; DB 7; Length 1070;
Best Local Similarity 23.2%; Pred. No. 42;
Matches 23; Conservative 20; Mismatches 47; Indels 9; Gaps 2;

Qy 37 EGKKDAGVYVNLKDTF-----IKPVFKIEEKEEENKPTFDVSKKKNPQVNHSQLN 90
      :||| : : : ||| : ||| : |||
Db 586 DGDSDKKVMEVGKSSDSGSVEMKPTAESLIEDVKDENASKTVDVKQETGSPDTKKKEGA 645
      :||| : : : ||| : ||| : |||

Qy 91 ESHRKEDLQREEH---SQKSDSTKDVDTATVLDKNNISSK 126
      :||| : : : ||| : ||| : |||
Db 646 SSSSKDQTKTGEDKKAEEKNNSETMTMSEGGKIDRNNITDEK 684
      :||| : : : ||| : ||| : |||

RESULT 21
US-11-096-568A-28313
; Sequence 28313, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28313
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1276)
; OTHER INFORMATION: Ceres Seq. ID no. 2712008
US-11-096-568A-28313

Query Match 11.4%; Score 79.5; DB 7; Length 1276;
Best Local Similarity 23.2%; Pred. No. 52;
Matches 23; Conservative 20; Mismatches 47; Indels 9; Gaps 2;

Qy 37 EGKKDAGVYVNLKDTF-----IKPVFKIEEKEEENKPTFDVSKKKNPQVNHSQLN 90
      :||| : : : ||| : ||| : |||
Db 792 DGDSDKKVMEVGKSSDSGSVEMKPTAESLIEDVKDENASKTVDVKQETGSPDTKKKEGA 851
      :||| : : : ||| : ||| : |||

Qy 91 ESHRKEDLQREEH---SQKSDSTKDVDTATVLDKNNISSK 126
      :||| : : : ||| : ||| : |||
```



```

; : || : : : || : : : || : : : || : : || : || :
Db 81 IEKQILKPENLRFHMLKQDNPTLYNKEKKIHNLQIEDTNLQHSQIPSSHPHQLK 140

QY 100 RE-----EHSQKSDSTKQVATVLDKNNI 123
; : || : || : || : || : || : || : || : || :
Db 141 OKKOISSETKNIQKINTKEQVQKITSONDL 171

RESULT 26
US-11-196-475-74
; Sequence 74, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-74

Query Match 11.3%; Score 78.5; DB 7; Length 700;
Best Local Similarity 24.3%; Pred. No. 31;
Matches 27; Conservative 29; Mismatches 32; Indels 23; Gaps 5;

QY 46 INLSKOTITKPVFKIEKEENKP-TFDVSKKNDQVN-----HSQLNES----H 93
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 LDIQRDVTREKLQENINETNKEKNLPKPGDVSSPKVKQLQTKESLEDLQELKETGSDN 375
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 RKEDLQREHSQKSD-----STKQVATVLDKNNISKSTNNPNK 134
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 QKRREIKQIEIKSDEKLLSKDKDRAKDGKALDLDR-ELNSKASSKEKSK 425

RESULT 27
US-10-793-626-1780
; Sequence 1780, Application US/10793626
; Publication No. US2005025478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic

```

```

; OTHER INFORMATION: amino acid sequence
US-10-793-626-1780

Query Match 11.3%; Score 78.5; DB 6; Length 1155;
Best Local Similarity 27.3%; Pred. No. 56;
Matches 24; Conservative 12; Mismatches 27; Indels 25; Gaps 2;

QY 63 EKKEENKPTFDVSKKDN-----POVNHSQLNESHRKEDLQREHSQK 106
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 100 EVKVEAPTTSDDVSKPRANEAVTNSTKPKTTTAPTVNEESIAETPTSTTQDSTERN 159
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 107 SDSTKQVATVLDKNNISKSTNNPNK 134
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 160 NPSLKD-----NLNSSSTTSKSK 178

RESULT 28
US-10-860-601-5
; Sequence 5, Application US/10860601
; Publication No. US20050273871A1
; GENERAL INFORMATION:
; APPLICANT: Asano, Yoshihiro
; APPLICANT: Takashima, Seiji
; APPLICANT: Kitakaze, Masafumi
; TITLE OF INVENTION: Method for Diagnosing Arrhythmogenic Right Ventricular Dysplasia
; FILE REFERENCE: 2144.0210000
; CURRENT APPLICATION NUMBER: US/10/860,601
; CURRENT FILING DATE: 2004-06-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: RVAP27
US-10-860-601-5

Query Match 11.2%; Score 78; DB 6; Length 191;
Best Local Similarity 23.9%; Pred. No. 7.5;
Matches 27; Conservative 29; Mismatches 37; Indels 20; Gaps 4;

QY 4 SSTTVSEEDFILP-----VYKGLKGYQFDGW--DISGPEGKKDAGYVNLKSDTFIK 55
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 SSSSEDEEYVVEKVLDRVVKGVYLLWKVGFSEHNTWEPEK-----NLDCPELIS 64
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 PVFKKIEKEENKPTFDVSKKNDPQVNHSQLNESHRKEDLQREHSQKSD 108
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 EFMKKYKMKEGEN-----NKPKEKSESNNKSNFNSADDIKSKKKKREQSN 111

RESULT 29
US-10-485-517-239
; Sequence 239, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629W0
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 239
; LENGTH: 405
; TYPE: PRT

```



```
; ORGANISM: Staphylococcus aureus
US-10-485-517-239

Query Match      11.1%; Score 77.5; DB 6; Length 405;
Best Local Similarity 29.7%; Pred. No. 20;
Matches 22; Conservative 16; Mismatches 33; Indels 3; Gaps 2;

QY 59 KIEKKKEENKPTFDVSKKKNPNQVNSQLN-ESHKEDLQREHHSQKSDTKDVTATV 117
DB 51 KAAESTNKLNEATTSASDNQSSDKVDMQQLNQEDNTKNDQKEMVSSQGNETTSGNKL 110

QY 118 LKNNISSKSTNN 131
DB 111 IEKESV--QSTTGN 122

RESULT 30
US-10-873-528-191
; Sequence 191, Application US/10873528
; Publication No. US200502768141
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 191
; LENGTH: 627
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-191

Query Match      11.1%; Score 77.5; DB 6; Length 627;
Best Local Similarity 23.4%; Pred. No. 33;
Matches 30; Conservative 26; Mismatches 59; Indels 13; Gaps 4;

QY 20 GLEKGY--QPDGWEISGEGKKDAGYVNLNLSKDTFIKPVKKEEENKPT----- 72
DB 497 GAKQGWKKANDWTFYKTDGSRAGWIKDKDKWYFLKENGQLLVNGKTPGTYVDSGA 556

QY 73 --FDVSKKKN--POVNSQLNESHKEDLQREHHSQKSDTKDVTATVLDKNNISSK 126
DB 557 WLVDVSIKSAIKTTSHEIKESKEVVKOLENKETSQHSVTFSTSQDLTSSTSSQS 616

QY 127 STNNPNK 134
DB 617 ETSVKNSE 624

RESULT 31
US-10-793-626-2482
; Sequence 2482, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2482
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2482

Query Match      11.1%; Score 77; DB 6; Length 568;
Best Local Similarity 24.8%; Pred. No. 33;
Matches 32; Conservative 22; Mismatches 41; Indels 34; Gaps 6;

QY 3 MSSTIVSEEDFILPVYKGELEKGYQFDGWEISGEGKKDAGYVNLNLSKDTFIKPVFKKIE 62
DB 234 VNEKILNQEDVVPVSLISEL---YQTQ-TSISN-----TYVL-----TFSDEVIKALN 277

QY 63 EKKKEENKPTFDVSKKKNPNQVNSQLNESHKEDLQREHHSQKSDTKDVTATVLDKNN 122
DB 278 KKIENESTPLPERA-----VNHQVNLSSDSNEDRYBYDR-----YIELNT 319

QY 123 ISSKSTNN 131
DB 320 LKDSILTSN 328

RESULT 32
US-11-196-475-70
; Sequence 70, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; TITLE OF INVENTION: Burgdorferi
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-11-196-475-70

Query Match      11.1%; Score 77; DB 7; Length 663;
Best Local Similarity 31.3%; Pred. No. 40;
Matches 26; Conservative 20; Mismatches 29; Indels 8; Gaps 4;

QY 55 KPVFKKIEKKKEENKPTFDVSKKKNPNQVNSQLN-ESHKEDLQREHHSQKSDTKD 112
DB 378 REIEKQIEKKND-----ELPKNDHKALDLKQLNLSKASSKEKIEGEEEDKELDCKN 432

QY 113 V-TATVLDKNNISSKSTNNPNK 134
DB 433 LEPVSEADKVDKISKNNNEVSK 455

RESULT 33
US-11-196-475-78
```



```
Db 139 EKRRKEIDGWRKQKQIASEGAKENANFV-PLGGDWRDRVKKRAKKEAKTEPIRAAA- 196
Qy 78 KDNFQVNHSQLNESHKKE-DLQREHSQKS-----DSTKDVATATVLDKNNISSKSTT 129
Db 197 -----EQHKGEPDLSLSKGPSCWQQAVIDESTKQ-----VYGGNNLTSETTW 239
Qy 130 NNPKN 134
Db 240 DRPSK 244

RESULT 37
US-11-232-440-33
; Sequence 33, Application US/11232440
; Publication No. US20060068434A1
; GENERAL INFORMATION:
; APPLICANT: STOECKER, JAY
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING CANCER USING
; TITLE OF INVENTION: COMPONENTS OF THE U2 SPLICEOSOMAL PARTICLE
; FILE REFERENCE: MTP-031
; CURRENT APPLICATION NUMBER: US/11/232,440
; PRIOR FILING DATE: 2005-09-21
; PRIOR APPLICATION NUMBER: 60/612,310
; PRIOR FILING DATE: 2004-09-21
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: Patent In Ver. 3.3
; SEQ ID NO 33
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-232-440-33

Query Match 10.9%; Score 76; DB 7; Length 482;
Best Local Similarity 21.4%; Pred. No. 33;
Matches 27; Conservative 24; Mismatches 39; Indels 36; Gaps 5;

Qy 34 SGFEGKDGAGYVINI-SKDTFKVPFKKIEEK-----EENKPTFDVSK 77
Db 40 SGLSQKEE-----EEDTFIEE--QLSEKLLERQRHLHEWLLREKQAQEFRIKK 90
Qy 78 KDNFQVNH-----SQLNESHKEDLQREHSQKSDTKQVATVLD--KNNISSK 126
Db 91 EKBEAAKTRQEQERKLKEQWEEOQKREBEQKQKKEALQMLDQAELENG 150
Qy 127 STNNP 132
Db 151 TTQNP 156

RESULT 38
US-11-045-004-1651
; Sequence 1651, Application US/11045004
; Publication No. US20060078901A1
; GENERAL INFORMATION:
; APPLICANT: BUCHRIESER, CARMEN
; APPLICANT: FRANGEUL, LIONEL
; APPLICANT: COUVE, ELISABETH
; APPLICANT: RUSNIOK, CHRISTOPHE
; APPLICANT: FSIHI, HAFIDA
; APPLICANT: DEHOUX, PIERRE
; APPLICANT: DUSBURGET, OLIVIER
; APPLICANT: CHETOUANI, FARID
; APPLICANT: NEDJARI, HAFED
; APPLICANT: GLASER, PHILIPPE
; APPLICANT: KUNST, FRANCK
; APPLICANT: COSSART, PASCALE
; APPLICANT: DANIELS, JUSTIN
; APPLICANT: GOEBEL, WERNER
; APPLICANT: KREFT, JURGEN
; APPLICANT: KUHN, MICHAEL
; APPLICANT: NG, EVA
; APPLICANT: VAZQUEZ-BOLAND, ANTONIO
```

```
; APPLICANT: DOMINGUEZ-BERNAL, GUSTAVO
; APPLICANT: GARRIDO-GARCIA, PATRICIA
; APPLICANT: TIERREZ-MARTINEZ, ALBERTO
; APPLICANT: AMEND, ALEXANDRA
; APPLICANT: CHAKRABORTY, TRINAD
; APPLICANT: DOMANN, EUGEN
; APPLICANT: HAIN, THORSTEN
; APPLICANT: BERCHE, PATRICK
; APPLICANT: CHARBIT, ALAIN
; APPLICANT: DURANT, LIONEL
; APPLICANT: PEREZ-DIAZ, JOSE-CLAUDIO
; APPLICANT: BAQUERO, FERNANDO
; APPLICANT: GARCIA DEL PORTILLO, FRANCISCO
; APPLICANT: GOMEZ-LOPEZ, NURIA
; APPLICANT: MADUENIO, ENCARNA
; APPLICANT: PABLOS, BETRIZ DE
; APPLICANT: WEHLAND, JURGEN
; APPLICANT: KARST, UWE
; APPLICANT: ENTIAN, KARL-DIETER
; APPLICANT: HAUF, JORG
; APPLICANT: ROSE, MATTHIAS
; APPLICANT: VOSS, HAMUT
; TITLE OF INVENTION: LISTERIA MONOCYTOGENES GENOME, POLYPEPTIDES AND USES
; FILE REFERENCE: 05394.0018-02
; CURRENT APPLICATION NUMBER: US/11/045,004
; CURRENT FILING DATE: 2005-01-28
; PRIOR APPLICATION NUMBER: 10/637,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: 10/257,023
; PRIOR FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: PCT/FR01/01118
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: FR 00/04,629
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 2854
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 1651
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-11-045-004-1651

Query Match 10.9%; Score 76; DB 7; Length 502;
Best Local Similarity 25.5%; Pred. No. 35;
Matches 24; Conservative 17; Mismatches 25; Indels 28; Gaps 4;

Qy 10 BEDFILPVYKGE-LEKGYQFDGWEI-----SGFEGKKD-----AGY 44
Db 6 EEEFLAYIKKMEALEBALVYWDLRTCAPAKGMEGRSDVIGLSEEIFNMQTSEMAAF 65
Qy 45 VINLSKDTFKVPFKKIEKKREKPKTFDVSCK 78
Db 66 IAGLNQD----KENLSEITRKTLESQKTYDLNKK 96

RESULT 39
US-10-793-626-2700
; Sequence 2700, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2700
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-2700

Query Match 10.8%; Score 75.5; DB 6; Length 278;  
Best Local Similarity 23.7%; Pred. No. 19;  
Matches 28; Conservative 16; Mismatches 45; Indels 29; Gaps 4;  
  
QY 46 INLSKDTFKVPFKIEKKEE-----ENKPTFDVSKKKONPQVNHSQLNESHKED-- 97  
Db ||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||||  
98 INKKEETNNDDGIEKSSSDRTSTTNDENEATFLQKSPQDNTHLTBEVEKPSVSESN 157  
QY 98 -----LQREHSHQSDSTKD-----VTATVLDKNNISSK--STTNPNK 134  
Db :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

158 SSIDTAQPSHTTIINRESVQTSNVEDSHVSDPANSKIESNTSGKEENTIEQPNK 215  
  
RESULT 40  
US-10-793-626-398  
; Sequence 398, Application US/10793626  
; Publication No. US2005025478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P03480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 398  
; LENGTH: 404  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (404)  
; OTHER INFORMATION: variable amino acid  
US-10-793-626-398

Query Match 10.8%; Score 75.5; DB 6; Length 404;  
Best Local Similarity 29.0%; Pred. No. 30;  
Matches 27; Conservative 18; Mismatches 33; Indels 15; Gaps 5;  
  
QY 49 SKDTFKVPFKIEKKEEENKPTFDVSKKKDNQVN-----HSQLNESHKEDLQREE 102  
Db :||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
161 TKNSTSTSEFTYLNHSPKSSSEVPSAIFGTKKRPIENGVIPEHKELND---KEIVQODE 217  
  
QY 103 --HSQKS-DSTKDVAT---VLDKNNISSKSTT 129  
Db ||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
218 VSHSTKSIDASKNVNSNDNNIEKNQKKQQT 250

Search completed: April 24, 2006, 15:44:47  
Job time : 11.4922 secs

GenCore version 5.1.7  
Copyright (C) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 67.1499 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773  
Perfect score: 696  
Sequence: 1 KEMSTIVSEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVILNSKDTFIKPVFKK 134

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	696	100.0	773	4	US-10-067-385-8
2	696	100.0	2119	3	US-09-769-744A-28
3	696	100.0	2140	4	US-10-282-122A-73670
4	696	100.0	2140	5	US-10-472-928-1180
5	693	99.6	637	5	US-10-617-320-3169
6	615	88.4	117	3	US-09-765-272-68
7	615	88.4	117	6	US-11-106-649-68
8	107	15.4	188	5	US-10-691-672A-7
9	103.5	14.9	169	5	US-10-691-672A-2
10	103.5	14.9	647	5	US-10-691-672A-3
11	101.5	14.6	564	6	US-11-097-143-12723
12	99.5	14.3	707	4	US-10-282-122A-52942
13	98	14.1	665	3	US-09-820-843A-107
14	93	13.4	869	4	US-10-437-963-12282
15	93	13.4	1529	5	US-10-732-923-8762
16	93	13.4	3127	5	US-10-732-923-22588
17	92.5	13.3	2468	4	US-10-755-889-615
18	92.5	13.3	2468	5	US-10-489-740-216
19	92.5	13.3	2519	5	US-10-450-763-4695
20	92	13.2	1005	4	US-10-437-963-18765
21	91	13.1	470	5	US-10-739-930-6262
22	89.5	12.9	511	4	US-10-289-762-509
23	88.5	12.7	540	5	US-10-732-923-22820
24	88.5	12.7	903	4	US-10-282-122A-52328
25	88	12.6	225	4	US-10-032-585-7829
26	88	12.6	778	4	US-10-724-972A-5663
27	88	12.6	815	5	US-10-496-905-23

28	88	12.6	1980	5	US-10-482-834A-144	Sequence 144, Appl
29	87.5	12.6	1196	4	US-10-282-122A-52737	Sequence 52737, A
30	87.5	12.6	1702	3	US-09-839-996-5	Sequence 5, Appli
31	87.5	12.6	1702	4	US-10-080-505-5	Sequence 5, Appli
32	87.5	12.6	1702	4	US-10-645-655-5	Sequence 5, Appli
33	87.5	12.6	1702	4	US-10-687-046-5	Sequence 5, Appli
34	87.5	12.6	3124	5	US-10-732-923-22709	Sequence 22709, A
35	87	12.5	281	6	US-11-097-143-21276	Sequence 174814, A
36	87	12.5	593	4	US-10-424-599-174814	Sequence 52510, A
37	87	12.5	932	4	US-10-282-122A-52510	Sequence 16976, A
38	86.5	12.4	1373	5	US-10-732-923-16976	Sequence 149572, A
39	86	12.4	145	4	US-10-424-599-149572	Sequence 238086, A
40	86	12.4	835	4	US-10-425-115-238086	Sequence 187664, A
41	86	12.4	1075	5	US-10-942-522-6	Sequence 6, Appli
42	86	12.4	1726	5	US-10-474-778-6	Sequence 150, Appl
43	86	12.4	1980	4	US-10-482-834A-150	Sequence 8, Appli
44	86	12.4	1980	5	US-10-942-522-8	
45	86	12.4	1980	5	US-10-942-522-8	

## ALIGNMENTS

## RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 696;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 3.6e-54;	Mismatches 0;	Indels 0; Gaps 0;
Matches 134;	Conservative 0;			
Qy	1	KEMSTIVSEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVILNSKDTFIKPVFKK	60	
Db	640	KEMSTIVSEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVILNSKDTFIKPVFKK	699	
Qy	61	IEEKKBENKPTFDVSKKKONPQVNHSQLNESHKBDLQREHSQKSDTKDVTATVLDK	120	
Db	700	IEEKKBENKPTFDVSKKKONPQVNHSQLNESHKBDLQREHSQKSDTKDVTATVLDK	759	
Qy	121	NNISSKSTTNPNK	134	
Db	760	NNISSKSTTNPNK	773	

## RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins

```
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match      100.0%; Score 696; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 1.2e-53;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKGAGYVINLSKDTFIKPVFKK 60
    |||||||
Db 1952 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKGAGYVINLSKDTFIKPVFKK 2011
    |||||||

QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120
    |||||||
Db 2012 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2071
    |||||||

QY 121 NNISKSTTNPNK 134
    |||||||
Db 2072 NNISKSTTNPNK 2085
    |||||||

RESULT 3
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match      100.0%; Score 696; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.2e-53;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKGAGYVINLSKDTFIKPVFKK 60
    |||||||
Db 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKGAGYVINLSKDTFIKPVFKK 2032
    |||||||

QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120
    |||||||
Db 2033 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2092
    |||||||

QY 121 NNISKSTTNPNK 134
    |||||||
Db 2093 NNISKSTTNPNK 2106
    |||||||

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
US-10-472-928-1180

Query Match      100.0%; Score 696; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.2e-53;
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKGAGYVINLSKDTFIKPVFKK 60
    |||||||
Db 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKGAGYVINLSKDTFIKPVFKK 2032
    |||||||

QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120
    |||||||
Db 2033 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2092
    |||||||

QY 121 NNISKSTTNPNK 134
    |||||||
Db 2093 NNISKSTTNPNK 2106
    |||||||

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```

```
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneko
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-10-617-320-3169
Query Match 99.6%; Score 693; DB 5; Length 637;
Best Local Similarity 99.3%; Pred. No. 5.3e-54; Mismatches 0; Indels 0; Gaps 0;
Matches 133; Conservative 1;
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKIPVFKK 60
Db KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKIPVFKK 529
QY 61 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120
Db IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 589
QY 121 NNISKSTNNPNK 134
Db NNISKSTNNPNK 603
RESULT 6
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US2002006154A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68
Query Match 88.4%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.7e-48; Mismatches 0; Indels 0; Gaps 0;
Matches 117; Conservative 0;
QY 18 YKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKIPVFKKEKKKEENKPTFDVSK 77
Db YKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFKIPVFKKEKKKEENKPTFDVSK 60
QY 78 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTNNPNK 134
Db KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISKSTNNPNK 117
RESULT 7
US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
```

```
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      88.4%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.7e-48;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 YKLEKGYQPDGWEISFEGKDGAGYINLSKDTFIKVPFKKIEKKKEENKPTFDVSK 77
DB 1 YKLEKGYQPDGWEISFEGKDGAGYINLSKDTFIKVPFKKIEKKKEENKPTFDVSK 60
QY 78 KDNQPVNHSQLNESHKEDLQREHSHQSKSDSTKDVATVLDKNNISSKSTNNPNK 134
DB 61 KDNQPVNHSQLNESHKEDLQREHSHQSKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 8
US-10-691-672A-7
; Sequence 7, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 7
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; LOCATION: (1)..(188)
; OTHER INFORMATION: MSP3a to MSP3f
US-10-691-672A-7

Query Match      15.4%; Score 107; DB 5; Length 188;
Best Local Similarity 23.8%; Pred. No. 0.1;
Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;

QY 1 KEMSTTVSBEEDFILPVYKGELEKGYQDGEWISGF--EGKDGAG-----YVINLSKDTF 53
DB 19 KEASS-----YDYIL-----GWFFGGVPEHKKEENMLSHLYVSSKDKENI 59
QY 54 IKPVFKKIEKKKEENKPTFDVSKKDNQPVNHSQLNESHKEDLQ 100
DB 60 SKENDDVLDKEKEEAETEEELKKEETETSEISEDEEEEEEEKEENDKKKEQEK 119
QY 101 EHSQKSDSTKDVATVLDKNNISSKSTNN 131
DB 120 EQSNENNDQKDMAE-----QNLISKQNNN 145

RESULT 9
US-10-691-672A-2
; Sequence 2, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 2
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(169)
; OTHER INFORMATION: MSP3 amino acids 212-380
US-10-691-672A-2

Query Match      14.9%; Score 103.5; DB 5; Length 169;
Best Local Similarity 25.2%; Pred. No. 0.19;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 1 KEMSTTVSBEEDFILPVYKGELEKGYQDGEWISGF--EGKDGAG-----YVINLSKDTF 53
DB 1 KEASS-----YDYIL-----GWFFGGVPEHKKEENMLSHLYVSSKDKENI 41
QY 54 IKPVFKKIEKKKEENKPTFDVSKKDNQPVNHSQLNESHKEDLQ 100
DB 42 SKENDDVLDKEKEEAETEEELKKEETETSEISEDEEEEEEEKEENDKKKEQEK 100
QY 101 EHSQKSDSTKDVATVLDKNNISSKSTNN 131
DB 101 EQSNENNDQKDMAE-----QNLISKQNNN 126

RESULT 10
US-10-691-672A-3
; Sequence 3, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 3
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(647)
; OTHER INFORMATION: GLURP MSP3 fusion protein
US-10-691-672A-3

Query Match      14.9%; Score 103.5; DB 5; Length 647;
Best Local Similarity 25.2%; Pred. No. 0.93;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 1 KEMSTTVSBEEDFILPVYKGELEKGYQDGEWISGF--EGKDGAG-----YVINLSKDTF 53
DB 479 KEASS-----YDYIL-----GWFFGGVPEHKKEENMLSHLYVSSKDKENI 519
QY 54 IKPVFKKIEKKKEENKPTFDVSKKDNQPVNHSQLNESHKEDLQ 100
DB 520 SKENDDVLDKEKEEAETEEELKKEETETSEISEDEEEEEEEKEENDKKKEQEK 578
QY 101 EHSQKSDSTKDVATVLDKNNISSKSTNN 131
DB 579 EQSNENNDQKDMAE-----QNLISKQNNN 604

RESULT 11
US-11-097-143-12723
; Sequence 12723, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
```



```

; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12723
; LENGTH: 564
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-12723

Query Match      14.6%; Score 101.5; DB 6; Length 564;
Best Local Similarity 24.5%; Pred. No. 1.2;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 11 EDFILPVVKGELKGYQFDG-----HISPEGKKDAGYVI-----NLSKDTPIK 55
DB 78 ELDLTFLESRSFK--VFDGWVDEHRDHGDHVDQPSGEALDDHDEHDDHDEHDEDEE 135
QY 56 PVFKKIEKKREKPT-----PDVSKKONPQVNHSQLNESHKRLQREHHSQKSDS 109
DB 136 PLTELELELEEEPTDEPADEYEDEDEENNA--GENITAEADAEEBEEEDND 193
QY 110 TKDVTATVLDKNISKST 128
DB 194 EGTVEATVEATTTEAT 212

RESULT 12
US-10-282-122A-52942
; Sequence 52942, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52942
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Clostridium difficile
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (6)..(6)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(18)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (29)..(29)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (37)..(37)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (43)..(43)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (54)..(54)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (84)..(84)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (86)..(86)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (359)..(359)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (385)..(385)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (388)..(388)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (396)..(396)
; OTHER INFORMATION: X=any amino acid
; FEATURE:

```

```
; NAME/KEY: MISC FEATURE
; LOCATION: (400)..(400)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (402)..(402)
; OTHER INFORMATION: X-any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (404)..(404)
; OTHER INFORMATION: X-any amino acid
US-10-282-122A-52942

Query Match      14.3%; Score 99.5; DB 4; Length 707;
Best Local Similarity 25.3%; Pred. No. 2.4;
Matches 37; Conservative 26; Mismatches 60; Indels 23; Gaps 4;

QY 8 VSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGYVYINLSKOTFIKPVFKKIEEK--- 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 ISIEDDAEAGVKEEIDSNQDIDGVV---EDKDTTDKEYDSNKEDIISPENKSKKKAKL 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 65 -----KEEN-----KPTFDVSKKONPOVNHSQLNESHKEDLQREHHSQKSD 108
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 FGFIKKNEEVQEEENLNDISFDILDKPVENNVQKSEIEQNELKE-IKQEEPSQHIE 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 109 SFKDVYATVLDKNNISSKSTTNNPK 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 EERSVKIEKPINNLDKVSNNESK 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-09-820-843A-107
; Sequence 107, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|3845248
US-09-820-843A-107

Query Match      14.1%; Score 98; DB 3; Length 665;
Best Local Similarity 27.1%; Pred. No. 3;
Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;

QY 23 EKGYPDGMWEI--SGFEGKKGAGYVYINLSKOTFIKPVFKKIEEKKE-----EENK 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 158 EKGKQ-----DISNSNAENKQD-----VKEGVKLEBEKKKEKISDDHKVBEKK 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 71 PTFD----VSKKDPVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISS 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 KSDDHKVEENKSKDDHKVBEENKSDDHKTEBVKVKEEHEEBE-----DKKKEKS 252
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 126 KSTTNNPK 134
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 ENKNKDNK 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-10-437-963-122282
```

```
; Sequence 122282, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122282
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MR_4530_25224C.1.pcp
US-10-437-963-122282

Query Match      13.4%; Score 93; DB 4; Length 869;
Best Local Similarity 45.1%; Pred. No. 12;
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;

QY 74 DVSKKONPOVNHSQLNESHKEDLQREHHSQKSDSTKDVTA--TVLDKNN 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 DASKRKNHQSEGNNL--SHRDEDPTRKRKKQKTNA'SDACAQBWVTEKNN 665
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-10-732-923-8762
; Sequence 8762, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8762
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1529)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-8762

Query Match      13.4%; Score 93; DB 5; Length 1529;
Best Local Similarity 27.2%; Pred. No. 23;
Matches 28; Conservative 19; Mismatches 30; Indels 26; Gaps 5;

QY 47 NLSKDT-----FIKPVFKKIEEKKEENKPTFDVSKKONPOVNHSQL-----NES 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 NIISDTQCIIKPIKYINSEYKKEEKKEEK-----NEKINTIHYSEISIKNSDNEQ 409
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 93 H-----RKEDLQREHHSQKSDSTKDVATVLDKNN--ISSKSTT 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 HPFLSKLRNVKKEKKKKKKTKIKTVIAQKNKTIVIAQKNKT 452
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 16
US-10-732-923-22588
; Sequence 22588, Application US/10732923
```





Db 49 EKSAFKESDFADLKESEK-----ALSDLSKLEBAVUN-----TLTKTKKES 96  
Qy 61 --IEEKKEENKPTFDVSKKONQVNHSQLNESHKEDLQREHSQKSDSTKQVATVL 118  
Db 97 SPKEKEEVVKEPBEVEKKE--EAAEKVEEKKSEAVVTEAPKAETVEAVVTEBII 154  
Qy 119 DKNISS 125  
Db 155 PKREVTT 161

## RESULT 24

US-10-282-122A-52328  
; Sequence 52328, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Lianguo  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 52328  
; LENGTH: 903  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-10-282-122A-52328

Query Match 12.7%; Score 88.5; DB 4; Length 903;  
Best Local Similarity 26.8%; Pred. No. 32;  
Matches 38; Conservative 22; Mismatches 47; Indels 35; Gaps 8;  
Qy 1 KEMSTIVSEDFILPVYKGELEKGYQDFGWEISGFEKQKAGYVNLKDTFIKPVFK 60  
Db 196 BEKSKQIIKEHNE-----KNEKSKIPKESIKL-----KBEKD-----KQVPE 235  
Qy 61 IEKKEEENKPTFDVSKKONQVNHSQLNESHKEDLQREHSQKSDSTKQVATVL 114  
Db 236 HNESEENKQISKVYNVELKNERSKQPKIN-VELKEEKNQSI--KEHNELEBETK--- 289

Qy 115 ATVLDKNNISSKSTNN--PNK 134  
Db 290 -KKLPKVNIELKSETKKQVFNK 310  
RESULT 25  
US-10-032-585-7829  
; Sequence 7829, Application US/10032585  
; Publication No. US20030180953A1  
; GENERAL INFORMATION:  
; APPLICANT: Terry, Roemer D.  
; APPLICANT: Bo, Jiang  
; APPLICANT: Charles, Boone  
; APPLICANT: Howard, Bussey  
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery  
; FILE REFERENCE: 10182-005-999  
; CURRENT APPLICATION NUMBER: US/10/032,585  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 8000  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 7829  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-10-032-585-7829

Query Match 12.6%; Score 88; DB 4; Length 225;  
Best Local Similarity 28.0%; Pred. No. 6.6;  
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;  
Qy 26 YQFQWEISGFEKQKAGYVNLKDTFIKPVFK-----IEKKEEENKPTFDVSKK-- 79  
Db 80 YDDDDDFEGFESSNGAAKELNLSQAIEKWKQRDLIEEREKLNKSKKEEIEKAKS 139  
Qy 80 --DNPQVNHSQLNESHKEDLQREH--SQKSDSTKQVATVLDKNN 122  
Db 140 TIDDFYENYNKRNHQKILSEQEKFISKDDFLK--RGLMDRVN 184

## RESULT 26

US-10-724-972A-5663  
; Sequence 5663, Application US/10724972A  
; Publication No. US20040147734A1  
; GENERAL INFORMATION:  
; APPLICANT: Bush, David  
; APPLICANT: Doucette-Stamm, Lynn  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: PATH03-16  
; CURRENT APPLICATION NUMBER: US/10/724,972A  
; CURRENT FILING DATE: 2003-12-01  
; PRIOR APPLICATION NUMBER: 09/450,969  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 7544  
; SEQ ID NO 5663  
; LENGTH: 778  
; TYPE: PRT  
; ORGANISM: S. epidermidis  
US-10-724-972A-5663

Query Match 12.6%; Score 88; DB 4; Length 778;  
Best Local Similarity 25.7%; Pred. No. 29;  
Matches 37; Conservative 21; Mismatches 50; Indels 36; Gaps 7;  
Qy 4 SSTIVSEDFILPVYKGELEKGYQDFGWEISGFE-----EISGFE-----GKQAGYVNL--LSKD 51

```
Db 643 SGTKVLPHSKVLMATDGLTMP-DMTGWTKEEDVLAFLDTIKVSTKGNGFVTNQSI SKG 701
Qy 52 TFIKVPFKIIEKKEEENKPTFDVS-----KKKDNQPNVHNSQLNESHKREDLQREHSOKS 107
Db 702 QIIK-----NKDKIEVLSAEDTDDQEKTDSDSNKSKKKADEHSNTS 748
Qy 108 DSTKDVATATVLDKNNISSKSTTN 131
Db 749 SSTKN-----DKSNADSKNDSDD 766

RESULT 27
US-10-496-905-23
; Sequence 23, Application US/10496905
; Publication No. US20050192215A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Malabika
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhao, Qing
; APPLICANT: Xu, Chongjun
; APPLICANT: Mulero, Julio J.
; APPLICANT: Boyle, Bryan J.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL POLYPEPTIDES AND POLYNUC
; FILE REFERENCE: HVS-BICIP/US
; CURRENT APPLICATION NUMBER: US/10/496,905
; CURRENT FILING DATE: 2004-05-26
; PRIOR FILING DATE: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: PCT/US00/34263
; PRIOR FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/598,075
; PRIOR FILING DATE: 2000-08-20
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 815
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-496-905-23

Query Match 12.6%; Score 88; DB 5; Length 815;
Best Local Similarity 27.3%; Pred. No. 31;
Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;

Qy 20 GELEKGYQDGEWISG--PEGKDGAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSK 77
Db 196 GOKKKQESFKSWEASGKHQEVSKPAVSLQQRKQDTSKLRLPPEQKKQEI SKSPSPSQ 255
Qy 78 -KKDNQPNVHNSQLNESHKRED-----LQREHSQKSDSTKVATVLDKNNISSKST 128
Db 256 WKQDTFKSKAGVQSEHKQETPKLWVPVLQKEQ-DPKKQTPKSWTPSMQSEQNTIKSWT 314
Qy 129 T 129
Db 315 T 315
```

```
RESULT 28
US-10-482-834A-144
; Sequence 144, Application US/10482834A
; Publication No. US20050074764A1
; GENERAL INFORMATION:
; APPLICANT: Mulley, John Charles
; APPLICANT: Harkin, Louise Anne
; APPLICANT: Dibbens, Michelle
; APPLICANT: Wallace, Robyn
; APPLICANT: Phillips, Hilary Anny
; APPLICANT: Heron, Sara Elizabeth
; APPLICANT: Berkovic, Samuel Frank
; APPLICANT: Scheffer, Ingrid Eileen
; APPLICANT: Bionomics Limited
; TITLE OF INVENTION: MUTATIONS IN ION CHANNELS
; FILE REFERENCE: 1386/17
; CURRENT APPLICATION NUMBER: US/10/482,834A
; CURRENT FILING DATE: 2004-01-02
; NUMBER OF SEQ ID NOS: 173
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 144
; LENGTH: 1980
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-482-834A-144

Query Match 12.6%; Score 88; DB 5; Length 1980;
Best Local Similarity 26.2%; Pred. No. 90;
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;

Qy 27 QPDGWEISGPEGKDGAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSKKKNQPNVH 86
Db 993 EMNQLQISVIRIKKGVAWT-KLVHAFMQAHFK---QREADEVKPLDELYEKKANCIANH 1048
Qy 87 SQLNESHKREDLQREHSQKSDSTKVATVLDKNNISSKSTTNPN 133
Db 1049 TSA-DIHRNGDFQKNGGTSGIGSSVEKYIIDEDHM---SFNNPN 1091

RESULT 29
US-10-282-122A-52737
; Sequence 52737, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 60/191,078
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
```

```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52737
; LENGTH: 1196
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52737

Query Match          12.6%; Score 87.5; DB 4; Length 1196;
Best Local Similarity 23.5%; Pred. No. 54;
Matches 40; Conservative 30; Mismatches 59; Indels 41; Gaps 7;

QY 1 KEMSTIVSEEDFILPVYKGE-----LEKGYQFDGWEISGFEGKDA----- 42
Db 693 KELEBTKTIEBFMGNILENKIKITLDEENLNKIDRIYNNIEITKPMGLNAIKETE 752
QY 43 --GYVINLSKO--TFIKPVFKKIEKKEENKPTFDVSKKONPVNHSQLNESHKEDL 98
Db 753 RLRSILNISREIKLTKDKIDIEENINESQQLBELAKKKD---LNHNDIKEC--EDFL 807
QY 99 QREHSQSDSTK---DVTATVLDKNNISSK-----STNNPNK 134
Db 808 QNEENVRNKKDLIEYKIEKAKLDEMLVSIKKELYSDMTNITNLNNENK 857

RESULT 30
US-09-839-996-5
; Sequence 5, Application US/09839996
; Publication No. US20030009010A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard P.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-839-996-5

Query Match          12.6%; Score 87.5; DB 3; Length 1702;
Best Local Similarity 27.2%; Pred. No. 83;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 46 INLSKDTPIKPVFKKIEKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHS- 104
Db 1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESPKRRRSI 1355
QY 105 --QKSDSTKQVTVATVLDKNNISSKSTNNPNK 134
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 31
US-10-080-505-5
; Sequence 5, Application US/10080505
; Publication No. US20030073166A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-080-505-5

Query Match          12.6%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 83;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 46 INLSKDTPIKPVFKKIEKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHS- 104
Db 1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESPKRRRSI 1355
QY 105 --QKSDSTKQVTVATVLDKNNISSKSTNNPNK 134
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 32
US-10-645-655-5
; Sequence 5, Application US/10645655
; Publication No. US20040063908A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
/
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/645,655
/ FILING DATE: 20-Aug-2003
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/296,791
/ FILING DATE: 25-AUG-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Trecartin, Richard F.
/ REGISTRATION NUMBER: 31,801
/ REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 781-1989
/ TELEFAX: (415) 398-3249
/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1702 amino acids
/ TYPE: amino acid
/ TOPOLOGY: unknown
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-645-655-5
Query Match 12.6%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 83;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 46 INLSKDTFTKPVFKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHS- 104
|| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1296 INTGSATLTETAESKDPQTETAASTEDASQHKANTVADNSVANNSSSEPKSRRSI 1355

QY 105 --QKSDSTKDVATVLDKNNISSKSTNNPNK 134
: : : : : : : : : : : : : : : : : : : : : : : :
DB 1356 SQPQTSABETTAASTDDETTIADNKRKSKPNR 1387

RESULT 33
US-10-687-046-5
; Sequence 5, Application US/10687046
; Publication No. US20040157241A1
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCR/DHR
; CURRENT APPLICATION NUMBER: US/10/687,046
; CURRENT FILING DATE: 2003-10-15
; PRIOR APPLICATION NUMBER: US/10/080,505
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-10-687-046-5

Query Match 12.6%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 83;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 46 INLSKDTFTKPVFKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHS- 104
|| : : : : : : : : : : : : : : : : : : : : : : : :
DB 1296 INTGSATLTETAESKDPQTETAASTEDASQHKANTVADNSVANNSSSEPKSRRSI 1355

QY 105 --QKSDSTKDVATVLDKNNISSKSTNNPNK 134
: : : : : : : : : : : : : : : : : : : : : : : :
DB 1356 SQPQTSABETTAASTDDETTIADNKRKSKPNR 1387

/
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/10/645,655
/ FILING DATE: 20-Aug-2003
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/296,791
/ FILING DATE: 25-AUG-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Trecartin, Richard F.
/ REGISTRATION NUMBER: 31,801
/ REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 781-1989
/ TELEFAX: (415) 398-3249
/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1702 amino acids
/ TYPE: amino acid
/ TOPOLOGY: unknown
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-645-655-5
Query Match 12.6%; Score 87.5; DB 5; Length 3124;
Best Local Similarity 24.0%; Pred. No. 1.7e+02;
Matches 43; Conservative 28; Mismatches 61; Indels 47; Gaps 6;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISG---FEGKKDAGYVINLSKDTFIKPV 57
| : : : : : : : : : : : : : : : : : : : : : : : :
DB 2000 KKEIKTIVSDDMFTSPVNIKEYNNEQERKKHIVGNLSYDKTKICPFIKFKGRIKK- 2058
|| : : : : : : : : : : : : : : : : : : : : : : : :
QY 58 FKIEKKKEENKPTF-----DVSKKKKNPQVNH 87
|| : : : : : : : : : : : : : : : : : : : : : : : :
DB 2059 -NKIEKKKEKYNNNFLYNDYSSYSPKYGDNNENFVIKIRKDFQKFDHFNFS 2117
|| : : : : : : : : : : : : : : : : : : : : : : : :
QY 88 QL-----NESHKKE---DLQREHSHQKSDSTKD-VTATVLDKNNISSKSTNNPNK 134
: : : : : : : : : : : : : : : : : : : : : : : :
DB 2118 KFLHYNPMKKNKKNKNNVNRNRYPNVTSSSKDGVSYNFLSDSLFSSDNEYSSDNE 2176
|| : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 35
US-11-097-143-21276
; Sequence 21276, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21276
; LENGTH: 281
; TYPE: PRT
```



[illegible]

Qy	27	OPDGM-----EISCFEKK-----DAGYVNLNKDPTFKPVFK-----KIEB 63
Db	453	EFADWVGTFQKEKVISISTGKGPVLLHYVYDSVILVMDKKNKFYSAPKGIYVKIRE 512
Qy	64	KXEEENKPTFDVSKKDKNPQNVHSQLNESH-RXEDLQREBHSQKSDSTKDVATVLDDKNN 122
Db	513	KQSANKNVTKQITSGNSNTSSNLKKNNNYYSKNKYLTTTNKNKENDNT-----QNN 563
Qy	123	ISSKSTNN 131
Db	564	INNNNNNN 572

```

RESULT 39
US-10-424-599-149572
; Sequence 149572, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 149572
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_106086C.1.pgp
US-10-424-599-149572

```

Query Match	12.4%;	Score 86;	DB 4;	Length 145;
Best Local Similarity	26.0%;	Pred. No. 5.9;		
Matches 33;	Conservative 24;	Mismatches 42;	Indels 28;	Gaps 6;
Qy	33	ISGPEGKKDAGYVNLKSDTPIKVFVKIKERK-KEE-----ENKPTF-----	DVSKK	78
Db	18	LSGYEERIRSYRLQSRVQAAEIIEMMLNKGPREIQSIIMOVLEKQATLQQAQAEKN		77
Qy	79	KDNQVNHQSQ--LNESHKED-----LQREHSQKSDSTKDTVATVLDKN--NIS		124
Db	78	ASKHQVNAETNMNSESMMEEETKLTMLKALQGVLEKADTESDVNAHQADSPNLS		137
Qy	125	SKSTNN	131	
Db	138	SASTTS	144	

```

RESULT 40
US-10-425-115-238086
; Sequence 238086, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 238086
; LENGTH: 835
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:

```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 20.8345 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTIVSEEDFILPVYKG.....ATVLDKNNISSKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgm2\_6/ptodata/1/1aa/5\_COMB.pep.\*
- 2: /cgm2\_6/ptodata/1/1aa/6\_COMB.pep.\*
- 3: /cgm2\_6/ptodata/1/1aa/H\_COMB.pep.\*
- 4: /cgm2\_6/ptodata/1/1aa/PCUS\_COMB.pep.\*
- 5: /cgm2\_6/ptodata/1/1aa/RE\_COMB.pep.\*
- 6: /cgm2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	696	100.0	773	2	US-09-590-991-8
2	693	99.6	637	2	US-09-107-433-3169
3	693	99.6	2138	2	US-09-583-110-5274
4	615	88.4	117	2	US-08-961-083-68
5	615	88.4	117	2	US-09-536-784-68
6	615	88.4	117	2	US-09-765-271-68
7	615	88.4	117	2	US-09-765-272A-68
8	101.5	14.6	347	2	US-09-248-796A-16224
9	95	13.6	348	2	US-09-538-092-1316
10	92.5	13.3	2468	2	US-09-976-594-726
11	92.5	13.3	2468	2	US-09-538-092-1135
12	92.5	13.3	2522	2	US-09-949-016-10237
13	89.5	12.9	511	2	US-09-198-452A-509
14	89.5	12.9	511	2	US-09-438-185A-475
15	88	12.6	746	2	US-09-710-279-652
16	88	12.6	778	2	US-09-134-001C-3868
17	87.5	12.6	1702	2	US-08-296-791-5
18	87.5	12.6	1702	2	US-09-839-986-5
19	87.5	12.6	1702	2	US-10-080-505-5
20	87.5	12.6	1702	2	US-10-645-655-5
21	87.5	12.6	1702	4	PCT-US95-10661A-5
22	86	12.4	243	2	US-09-248-796A-20306
23	86	12.4	280	2	US-09-248-796A-17646
24	86	12.4	1989	2	US-09-949-016-10076
25	85.5	12.3	109	2	US-09-248-796A-24668
26	85	12.2	465	2	US-09-134-001C-3856
27	85	12.2	472	2	US-09-710-279-658

28	85	12.2	720	2	US-09-710-279-2058	Sequence 2058, Ap
29	85	12.2	728	2	US-09-134-001C-4968	Sequence 4968, Ap
30	85	12.2	1976	2	US-09-024-020B-9	Sequence 9, Appli
31	85	12.2	1976	2	US-09-425-043-9	Sequence 9, Appli
32	85	12.2	1978	2	US-09-024-020B-3	Sequence 3, Appli
33	85	12.2	1978	2	US-09-425-043-3	Sequence 3, Appli
34	85	12.2	1988	2	US-09-024-020B-4	Sequence 4, Appli
35	85	12.2	1988	2	US-09-425-043-4	Sequence 4, Appli
36	84.5	12.1	654	2	US-10-172-502-10	Sequence 10, Appl
37	84	12.1	817	2	US-09-248-796A-20276	Sequence 20276, A
38	83.5	12.0	461	2	US-09-949-016-8508	Sequence 8508, Ap
39	83.5	12.0	476	2	US-09-248-796A-15008	Sequence 15008, A
40	83	11.9	278	2	US-09-949-016-10508	Sequence 10508, A
41	82.5	11.9	243	2	US-09-248-796A-21334	Sequence 21334, A
42	82	11.8	743	2	US-08-910-925-3	Sequence 3, Appli
43	82	11.8	743	2	US-09-949-016-6261	Sequence 6261, Ap
44	82	11.8	1024	2	US-09-270-767-44973	Sequence 44973, A
45	81.5	11.7	477	2	US-09-075-375F-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1993-08-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	Best Local Similarity	100.0%;	Score 696;	DB 2;	Length 773;
Matches 134;	Conservative	0;	Mismatches	0;	Gaps 0;
Qy	1	KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSKDTFIKPVFKK	60		
Db	640	KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKDGAGYVINLSKDTFIKPVFKK	699		
Qy	61	IEEKKKEENKPTFVSKKKNQPNVHSQLNESHKEDLOREHSQKSDTKDVTATVLDK	120		
Db	700	IEEKKKEENKPTFVSKKKNQPNVHSQLNESHKEDLOREHSQKSDTKDVTATVLDK	759		
Qy	121	NNISSKSTTNNPNK	134		
Db	760	NNISSKSTTNNPNK	773		

RESULT 2

US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...637
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-433-3169

Query Match          99.6%; Score 693; DB 2; Length 637;
Best Local Similarity 99.3%; Pred. No. 1.1e-66;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKK 60
Db 1 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKK 529

QY 61 IEKKEEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDK 120
Db 1 IEKKEEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDK 589

QY 121 NNISKSSTNNPNK 134
Db 1 NNISKSSTNNPNK 603

RESULT 3
US-09-583-110-5274
; Sequence 5274, Application US/09583110
; Patent No. 6699703
;
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
```

```
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5274
; LENGTH: 2138
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5274

Query Match          99.6%; Score 693; DB 2; Length 2138;
Best Local Similarity 99.3%; Pred. No. 5.7e-66;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKK 60
Db 1 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKK 2030

QY 61 IEKKEEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDK 120
Db 1 IEKKEEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDK 2090

QY 121 NNISKSSTNNPNK 134
Db 1 NNISKSSTNNPNK 2104

RESULT 4
US-08-961-083-68
; Sequence 68, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-961-083-68

Query Match          88.4%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.5e-59;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 YKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 77
```

```
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIEEKEENKPTFDVSK 60
Qy 78 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 134
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 5
US-09-536-784-68
; Sequence 68, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match 88.4%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.5e-59;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIEEKEENKPTFDVSK 77
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIEEKEENKPTFDVSK 60

Qy 78 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 134
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 6
US-09-536-784-68
; Sequence 68, Application US/09765271
; Patent No. 6887663
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,271
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/536,784
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-271-68

Query Match 88.4%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.5e-59;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIEEKEENKPTFDVSK 77
Db 1 YKGELEKGYQFDGWEISGFEKGDAGYVNLNLSKDTFKPVFKIEEKEENKPTFDVSK 60

Qy 78 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 134
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 7
US-09-765-272A-68
; Sequence 68, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/09/765,272A
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin J. Hymel
; REGISTRATION NUMBER: 45,414
; REFERENCE/DOCKET NUMBER: PB340P2C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 610-5790
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272A-68

Query Match      88.4%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.5e-59;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 YKLEKGYQPDGWEISGFEKGDAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSK 77
DB 1 YKLEKGYQPDGWEISGFEKGDAGYVNLKDTFIKPVFKIIEKKEEENKPTFDVSK 60

QY 78 KDNQVNHSQLNESHKEDLQREHHSOKSDSTKDVATVLDKNNISSKTTNNPNK 134
DB 61 KDNQVNHSQLNESHKEDLQREHHSOKSDSTKDVATVLDKNNISSKTTNNPNK 117

RESULT 8
US-09-248-796A-16224
; Sequence 16224, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16224
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16224

Query Match      14.6%; Score 101.5; DB 2; Length 347;
Best Local Similarity 21.4%; Pred. No. 0.0083;
Matches 31; Conservative 32; Mismatches 63; Indels 19; Gaps 4;

QY 2 EMSSTIVSEEDFILPVYKLEKGYQPDGWEISGF-----EGKKDAGYVNLKDTFIK 55
DB 202 KLSPLMIDEIDTPIRYND--KKWYVATSSLQNTVQTDLESSESEIGWEDDLEENYRTG 259

QY 56 PVFKKIEKKEENKPTFDVSKKDNQVNHSQLN-----ESHKEDLQREHHSOKSDS 109
DB 260 PVFKILDQLRE-----WKAKEQNPKKEEENLNQKPVAKQKQPNSTKKQKQTKQKQ 314

QY 110 TKDVTATVLDKNNISSKTTNNPNK 134
DB 315 TKKITPKTSKRMLEIGISTSTNNIK 339
```

## RESULT 9

```
US-09-538-092-1316
; Sequence 1316, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/539,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1316
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q14093
US-09-538-092-1316
```

```
Query Match      13.6%; Score 95; DB 2; Length 348;
Best Local Similarity 30.9%; Pred. No. 0.042;
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;
```

```
QY 19 KGELEKGYQPDGWEISGFEKGDAGYVNLKDTFIKPVFKIIEKKEEENKPTF---DV 75
DB 205 ESEGEKG---GTEKDSKKGKDS-----KKGKOSAIELQAVKADEKKDGGKDKANKGDE 256

QY 76 SK--KDNQVNHSQLN-----ESHKEDLQREHHSOKSDSTKD---VTATVLDKNNI 123
DB 257 SKDAKQAKELKKGKDKKXPSSTSDSKDDVKKE---SKDQATKDAKVAKKOTEKESA 313

QY 124 SSK 126
DB 314 DSK 316
```

## RESULT 10

```
US-09-976-594-726
; Sequence 726, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 726
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1
US-09-976-594-726
```

```
Query Match      13.3%; Score 92.5; DB 2; Length 2468;
Best Local Similarity 31.0%; Pred. No. 1.1;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;
```



```

US-09-134-001C-3868
; Sequence 3868, Application US/09134001C
; Patent NO. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GPC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3868
; LENGTH: 778

```



```
RESULT 18
US-09-839-996-5
; Sequence 5, Application US/09839996
; Patent No. 6642371
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/839,996
; FILING DATE: 20-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-839-996-5
Query Match 12.6%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 2.3;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 46 INLSKDTFIKPVFKIEKKEENKPTFDVSKKONQVNHSQLNESHKEDLOREHS- 104
Db 1296 INTGSAITAITAEKSKDPQTETAATEDASQHKANTVADNSVANNSESSEPKRRRSI 1355

Qy 105 --QKSDSTKDVATVLDKNNISSKSTNNPNK 134
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 19
US-10-645-655-5
; Sequence 5, Application US/10645655
; Patent No. 6815182
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/645,655
; FILING DATE: 20-Aug-2003
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Treccartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RPT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-645-655-5
Query Match 12.6%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 2.3;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy 46 INLSKDTFIKPVFKIEKKEENKPTFDVSKKONQVNHSQLNESHKEDLOREHS- 104
Db 1296 INTGSAITAITAEKSKDPQTETAATEDASQHKANTVADNSVANNSESSEPKRRRSI 1355

Qy 105 --QKSDSTKDVATVLDKNNISSKSTNNPNK 134
Db 1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 19
US-10-080-505-5
; Sequence 5, Application US/10080505
; Patent No. 6676948
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph W.
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
```

Db 1296 INTGATALTETAEKSDKPTQETAASTEDASQHKANTVADNSVANNSSSEPKSRRRRSI 1355  
QY 105 --QKSDSKDVTATVLDKNNISSKSTNNPNK 134  
Db 1356 SQPQTSABETTAAGTDETTIADNSKRSKPNR 1387

## RESULT 21

PCT-US95-10661A-5

; Sequence 5, Application PC/TUS9510661A

; GENERAL INFORMATION:

; APPLICANT: Washington University, et al.

; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSES: Flehr, Hobbach, Test, Albritton &amp; Herbert

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/10661A

; FILING DATE: 16-AUG-1995

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/296,791

; FILING DATE: 25-AUG-1994

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Trecartin, Richard F.

; REGISTRATION NUMBER: 31,801

; REFERENCE/DOCKET NUMBER: FP-59941/RFT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1702 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

PCT-US95-10661A-5

Query Match 12.6%; Score 87.5; DB 4; Length 1702;  
Best Local Similarity 27.2%; Pred. No. 2.3;  
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY 46 INLSKDTIKPVFKKIEKEEENKPTFDVSKKDNQPNVHNSQLNESHKEDLOREHS- 104

Db 1296 INTGATALTETAEKSDKPTQETAASTEDASQHKANTVADNSVANNSSSEPKSRRRRSI 1355

QY 105 --QKSDSKDVTATVLDKNNISSKSTNNPNK 134

Db 1356 SQPQTSABETTAAGTDETTIADNSKRSKPNR 1387

## RESULT 22

US-09-248-796A-20306

; Sequence 20306, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 20306  
; LENGTH: 243  
; TYPE: PRT  
; ORGANISM: Candida albicans  
; US-09-248-796A-20306

Query Match 12.4%; Score 86; DB 2; Length 243;  
Best Local Similarity 28.0%; Pred. No. 0.25;  
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;

QY 26 YQFDGWEISGPEGKDGAGYVNLKCTFIKPVFKK----IEKKKEENKPTFDVSKKK-- 79

Db 98 YDDDDDEFEGFESSGAKEINLSQIAKEWKQRRDLIEEREKLNKSKKEEIEIAKAS 157

QY 80 --DNPOVNHSQLNESHKEDLOREH--SQKSDSTKDVATVLDKNN 122

Db 158 TIDDFYNYNSKRDNHQKILLQEKFISKRDDFLK--RGLWDRVN 202

## RESULT 23

US-09-248-796A-17646

; Sequence 17646, Application US/09248796A

; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 17646

; LENGTH: 280

; TYPE: PRT

; ORGANISM: Candida albicans

; US-09-248-796A-17646

Query Match 12.4%; Score 86; DB 2; Length 280;  
Best Local Similarity 26.1%; Pred. No. 0.3;  
Matches 37; Conservative 23; Mismatches 50; Indels 32; Gaps 8;

QY 3 MSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDGAGYVNLKDTFIKPVFKKIE 62

Db 84 VASTFCSKYDPNVSFASNLGLF-----ELYSANKK-----KNSP--PSFEHHE 127

QY 63 -EKKEENKPTFDVSKKDNQPNVHNS--QLNESHKEDLOREHS-----QKSDSTK 111

Db 128 IHSSEENK----YLKHPQLQRHRLHNLHORVPKSHKYEKNGRTIINPIQLNDNVY 183

QY 112 DVTATVLDKNNISSKSTNNPN 133

Db 184 HINPTLLSSNG-STSTTTNNEN 204

## RESULT 24

US-09-949-016-10076

; Sequence 10076, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 1999-02-12

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10076
; LENGTH: 1989
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10076

Query Match      12.4%; Score 86; DB 2; Length 1989;
Best Local Similarity 26.2%; Pred. No. 4.1;
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;

QY 27 QPDGWEISGFGKQAGYVINISKDTPFKVPKKEEKKBERNKPTFDVSKKQNPQVNH 86
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1002 ENNLQISVIRIKKGAVMT-KLKVHAFMQAHFK--QREADEVKPLDELYEKKANCIA 1057
QY 87 SOLNESHREKDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPN 133
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1058 TGA-DIHRNGDFQKNGGTGTSIGSSVEKYIIDEDHM---SPINPN 1100

RESULT 25
US-09-248-796A-24668
; Sequence 24668, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 24668
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-24668

Query Match      12.3%; Score 85.5; DB 2; Length 109;
Best Local Similarity 32.9%; Pred. No. 0.096;
Matches 27; Conservative 14; Mismatches 26; Indels 15; Gaps 4;

QY 62 EKKKEEN-----KPTFDVSK--KQNPQVNHSQLNESHKEDLQREHS-QKSDSTK 111
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11 DDDEEENTKQNGKEDIDRNKQSKEDNSNANSTQAVKNLTETLKENHSDEKSDPTK 70
QY 112 DVTATVLDKNNISSKSTNNPN 133
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 ENS-----KQGVSKENTNAN 87

RESULT 26
US-09-134-001C-3856
; Sequence 3856, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3856
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3856

Query Match      12.2%; Score 85; DB 2; Length 465;
Best Local Similarity 32.6%; Pred. No. 0.76;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 61 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQ-----REEHSQKSDS--- 109
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 MEENKQNPKE--NMSNKDDNA---THLNDSHRNEDLELFRNKNNAQRERRRIDNOSK 56
QY 110 TKDVTAT-----VLDKNNISSKSTNNPNK 134
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 EKDATSQSQLETKPMDKFLDNHKS--HNQNK 86

RESULT 27
US-09-710-279-658
; Sequence 658, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-658

Query Match      12.2%; Score 85; DB 2; Length 472;
Best Local Similarity 32.6%; Pred. No. 0.77;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

QY 61 IEKKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQ-----REEHSQKSDS--- 109
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 MEENKQNPKE--NMSNKDDNA---THLNDSHRNEDLELFRNKNNAQRERRRIDNOSK 56
QY 110 TKDVTAT-----VLDKNNISSKSTNNPNK 134
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 EKDATSQSQLETKPMDKFLDNHKS--HNQNK 86

RESULT 28
US-09-710-279-2058
; Sequence 2058, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
```



; APPLICATION NUMBER: US/09/425,043  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/024,020  
; FILING DATE: 16-FEB-1998  
; APPLICATION NUMBER: US 60/039,447  
; FILING DATE: 26-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CLARK, JANET P.  
; REGISTRATION NUMBER: 34,799  
; REFERENCE/DOCKET NUMBER: R0020B-REG  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 852-3097  
; TELEFAX: (650) 855-5322  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1976 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-425-043-9

Query Match 12.2%; Score 85; DB 2; Length 1976;

Best Local Similarity 25.2%; Pred. No. 5.3;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

Qy 27 QFDGWEISGFGKDGAVINLSKDTFKPVFKIEKKKEENKPTDVSKKDNPQVNH 86  
Db 989 EMNNLQISVIRIKKGAVT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKNCIANH 1044  
Qy 87 SQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPN 133  
Db 1045 TGV-DIHRNGDFQKNGTTSIGSSVEKYIIDEDHM---SFINNPN 1087

RESULT 32

US-09-024-020B-3

; Sequence 3, Application US/09024020B

; Patent No. 6030810

; GENERAL INFORMATION:

; APPLICANT: DELGADO, STEPHEN G.

; APPLICANT: DIETRICH, PAUL S.

; APPLICANT: FISH, LINDA M.

; APPLICANT: HERMAN, RONALD C.

; APPLICANT: SANGAMESWARAN, LAKSHMI

; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE

; TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: JANET PAULINE CLARK

; STREET: 3401 HILLVIEW AVENUE, MS A2-250

; CITY: PALO ALTO

; STATE: CA

; COUNTRY: U.S.A.

; ZIP: 94304-1397

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/024,020B

; FILING DATE: 16-FEB-1998

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/039,447

; FILING DATE: 26-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: CLARK, JANET P.

; REGISTRATION NUMBER: 34,799

; REFERENCE/DOCKET NUMBER: R0020B-REG

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 852-3097  
; TELEFAX: (650) 855-5322  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1978 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-024-020B-3

Query Match 12.2%; Score 85; DB 2; Length 1978;

Best Local Similarity 25.2%; Pred. No. 5.3;

Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;

Qy 27 QFDGWEISGFGKDGAVINLSKDTFKPVFKIEKKKEENKPTDVSKKDNPQVNH 86  
Db 991 EMNNLQISVIRIKKGAVT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKNCIANH 1046  
Qy 87 SQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPN 133  
Db 1047 TGV-DIHRNGDFQKNGTTSIGSSVEKYIIDEDHM---SFINNPN 1089

RESULT 33

US-09-425-043-3

; Sequence 3, Application US/09425043

; Patent No. 6335172

; GENERAL INFORMATION:

; APPLICANT: DELGADO, STEPHEN G.

; APPLICANT: DIETRICH, PAUL S.

; APPLICANT: FISH, LINDA M.

; APPLICANT: HERMAN, RONALD C.

; APPLICANT: SANGAMESWARAN, LAKSHMI

; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE

; TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF

; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: JANET PAULINE CLARK

; STREET: 3401 HILLVIEW AVENUE, MS A2-250

; CITY: PALO ALTO

; STATE: CA

; COUNTRY: U.S.A.

; ZIP: 94304-1397

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/425,043

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 09/024,020

; FILING DATE: 16-FEB-1998

; APPLICATION NUMBER: US 60/039,447

; FILING DATE: 26-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: CLARK, JANET P.

; REGISTRATION NUMBER: 34,799

; REFERENCE/DOCKET NUMBER: R0020B-REG

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 852-3097

; TELEFAX: (650) 855-5322

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1978 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-09-425-043-3

Query Match 12.2%; Score 85; DB 2; Length 1978;  
Best Local Similarity 25.2%; Pred. No. 5.3;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;  
Qy 27 QFDGWEISGFEKGDAGYVNLSDTKTFIKVPFKTEEKKEENKPTFDVSKKONPQVNH 86  
Db 991 ENNNLQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1046  
Qy 87 SOLNESHKEDLQREHSHKSDSTKDVATVLDKNNISSKSTNNPN 133  
Db 1047 TGV-DIHRNGDFQKNGTTSIGSSVEKYIIDEDHM---SFINNPN 1089

RESULT 34

US-09-024-020B-4  
; Sequence 4, Application US/09024020B  
; Patent No. 6030810  
; GENERAL INFORMATION:  
; APPLICANT: DELGADO, STEPHEN G.  
; APPLICANT: DIETRICH, PAUL S.  
; APPLICANT: FISH, LINDA M.  
; APPLICANT: HERMAN, RONALD C.  
; APPLICANT: SANGAMESWARAN, LAKSHMI  
; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
; TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JANET PAULINE CLARK  
; STREET: 3401 HILLVIEW AVENUE, MS A2-250  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 94304-1397  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/09/024,020B  
; FILING DATE: 16-FEB-1998  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/039,447  
; FILING DATE: 26-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CLARK, JANET P.  
; REGISTRATION NUMBER: 34,799  
; REFERENCE/DOCKET NUMBER: R0020B-REG  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 852-3097  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1988 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-024-020B-4

Query Match 12.2%; Score 85; DB 2; Length 1988;  
Best Local Similarity 25.2%; Pred. No. 5.3;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;  
Qy 27 QFDGWEISGFEKGDAGYVNLSDTKTFIKVPFKTEEKKEENKPTFDVSKKONPQVNH 86  
Db 1001 ENNNLQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1056  
Qy 87 SOLNESHKEDLQREHSHKSDSTKDVATVLDKNNISSKSTNNPN 133

Db 1057 TGV-DIHRNGDFQKNGTTSIGSSVEKYIIDEDHM---SFINNPN 1099

RESULT 35

US-09-425-043-4  
; Sequence 4, Application US/09425043  
; Patent No. 6335172  
; GENERAL INFORMATION:  
; APPLICANT: DELGADO, STEPHEN G.  
; APPLICANT: DIETRICH, PAUL S.  
; APPLICANT: FISH, LINDA M.  
; APPLICANT: HERMAN, RONALD C.  
; APPLICANT: SANGAMESWARAN, LAKSHMI  
; TITLE OF INVENTION: NOVEL CLONED TETRODOTOXIN-SENSITIVE  
; TITLE OF INVENTION: SODIUM CHANNEL I-SUBUNIT AND A SPLICE VARIANT THEREOF  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: JANET PAULINE CLARK  
; STREET: 3401 HILLVIEW AVENUE, MS A2-250  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 94304-1397  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION NUMBER: US/09/425,043  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/024,020  
; FILING DATE: 16-FEB-1998  
; APPLICATION NUMBER: US 60/039,447  
; FILING DATE: 26-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CLARK, JANET P.  
; REGISTRATION NUMBER: 34,799  
; REFERENCE/DOCKET NUMBER: R0020B-REG  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 852-3097  
; TELEFAX: (650) 855-5322  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1988 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-425-043-4

Query Match 12.2%; Score 85; DB 2; Length 1988;  
Best Local Similarity 25.2%; Pred. No. 5.3;  
Matches 27; Conservative 25; Mismatches 47; Indels 8; Gaps 4;  
Qy 27 QFDGWEISGFEKGDAGYVNLSDTKTFIKVPFKTEEKKEENKPTFDVSKKONPQVNH 86  
Db 1001 ENNNLQISVIRIKKGVAWT-KVKVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1056  
Qy 87 SOLNESHKEDLQREHSHKSDSTKDVATVLDKNNISSKSTNNPN 133

Db 1057 TGV-DIHRNGDFQKNGTTSIGSSVEKYIIDEDHM---SFINNPN 1099

RESULT 36

US-10-172-502-10  
; Sequence 10, Application US/10172502  
; Patent No. 6841154  
; GENERAL INFORMATION:  
; APPLICANT: FOSTER, Timothy et al.  
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES. . .

FILE REFERENCE: P07263US01/BAS  
CURRENT APPLICATION NUMBER: US/10/172,502  
PRIOR FILING DATE: 2002-06-17  
PRIOR APPLICATION NUMBER: US 60/298,098  
PRIOR FILING DATE: 2001-06-15  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 10  
LENGTH: 654  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-10-172-502-10

Query Match 12.1%; Score 84.5; DB 2; Length 654;  
Best Local Similarity 28.6%; Pred. No. 1.4;  
Matches 30; Conservative 18; Mismatches 44; Indels 13; Gaps 5;  
QY 41 DAGVYINL-SKDTETKPVFKIEEKEENKPTPDV-----SKKDNQVNHSQLNESHR 94  
DB 450 DQYHVRIVDKAFTKANTDKSNKKEQDQNSAKKEATPATPSKPTPSVPEKESQKQDSQK 509  
QY 95 KEDLQ-----REHSQKSDSTKDVATVLDKNNISSKSTNNPNK 134  
DB 510 DDNKQLPSVEKENDASSGKDKTPTATPTKGEVSSSTT--PTK 552

RESULT 37  
US-09-248-796A-20276  
Sequence 20276, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 20276  
LENGTH: 817  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-20276

Query Match 12.1%; Score 84; DB 2; Length 817;  
Best Local Similarity 27.0%; Pred. No. 2.1;  
Matches 24; Conservative 17; Mismatches 34; Indels 14; Gaps 4;  
QY 43 GVINLSK-----DTFIKPVFKIEEKEENKPTPDVSKKDNQVNHSQLNESHRKED 97  
DB 312 GLTISQSLDKLASTVQPILDIEGKAA--KRQIDIEKKQ-----KELELQLHEK-- 362  
QY 98 LQREHSQKSDSTKDVATVLDKNNISSK 126  
DB 363 AKKEEHEAKEKEKIDIBIAKLERNQNDK 391

RESULT 38  
US-09-949-016-8508  
Sequence 8508, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8508  
LENGTH: 461  
TYPE: PRT  
ORGANISM: Human  
US-09-949-016-8508

Query Match 12.0%; Score 83.5; DB 2; Length 461;  
Best Local Similarity 21.7%; Pred. No. 1.1;  
Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4;  
QY 40 KDAGVYINLSKDTETKPVFKIEEKK-----EENKPTPDVSKKKNPQ 83  
DB 26 RDSGLSQEEEDTPIEB--QQLEEEKLLERERLHBEWLLREQKAQEEFRIKKEKEAA 83  
QY 84 VN-----HSQLESHERKEDLQREHSQKSDSTKDVATVLD--KNNISSKSTNNP 132  
DB 84 KMWLEQERKLEQWKQQREREREKQKEKEKEAQQMLDQENDLENSTTQNP 143

RESULT 39  
US-09-248-796A-15008  
Sequence 15008, Application US/09248796A  
Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 15008  
LENGTH: 476  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-15008

Query Match 12.0%; Score 83.5; DB 2; Length 476;  
Best Local Similarity 27.6%; Pred. No. 1.1;  
Matches 27; Conservative 19; Mismatches 27; Indels 25; Gaps 3;  
QY 48 LSKDTETKPVFKIEEKEENKPTPDVSKKKNPQVNH-----SQL 89  
DB 181 LRKDMKAK--LKLKQKCKFNDELTKDLFKKOOTPEASNPFGNSNPFNGLNPFSSK 237  
QY 90 NESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKS 127  
DB 238 PEEKEKEKEKETSKSVADVAS-----KNAPFKPS 271

RESULT 40  
US-09-949-016-10508  
Sequence 10508, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:40:21 ; Search time 84.0872 Seconds  
(without alignments)  
700.187 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTVSEEDFILPVYKG.....ATVLDKNKISSKSTNNPNK 134

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_21:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*
- 9: Geneseq2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	696	100.0	773	4	AAB48343
2	696	100.0	2120	3	AA81710
3	696	100.0	2140	6	ABU01020
4	696	100.0	2140	6	ABU45746
5	696	100.0	2140	8	ADM92113
6	696	100.0	2140	8	ADT50099
7	693	99.6	637	8	ADR94534
8	693	99.6	637	9	AEA58404
9	693	99.6	2138	8	ADK48759
10	615	88.4	117	2	AAW55096
11	615	88.4	117	5	ABF54590
12	615	88.4	117	7	ADC45149
13	107	15.4	188	9	ADZ79639
14	107	15.4	354	9	ADZ72253
15	103.5	14.9	169	9	ADZ79634
16	103.5	14.9	647	9	ADZ79635
17	103.5	14.9	651	8	ADOL9012
18	103.5	14.9	651	8	ADOL9010
19	101.5	14.6	564	4	ABB61977
20	99.5	14.3	707	6	ABU25018
21	98	14.1	665	3	ABE18278
22	98	14.1	665	7	ABO23606
23	93	13.4	1791	8	ADP25441
24	92.5	13.3	2468	6	ABR64281

25	92.5	13.3	2468	7	ADE62723	Ade62723 Human Pro
26	92.5	13.3	2468	7	ADE62719	Ade62719 Human Pro
27	92.5	13.3	2468	7	ADE62727	Ade62727 Human Pro
28	92.5	13.3	2468	7	ADE62715	Ade62715 Human Pro
29	92.5	13.3	2468	8	ADL12997	Adl12997 Human ste
30	92.5	13.3	2468	8	ADN05260	Adn05260 Antipsori
31	92.5	13.3	2468	8	ADR14614	Adr14614 Human NF-
32	92.5	13.3	2519	4	ABG16636	Abg16636 Novel hum
33	92.5	13.3	2527	8	ADN04561	Adn04561 Antipsori
34	91.5	13.1	635	8	ADS93954	AdS93954 Fibrinoge
35	91.5	13.1	635	8	ADV83292	Adv83292 Streptoco
36	91.5	13.1	643	8	ADV89902	Adv89902 Streptoco
37	91.5	13.1	643	8	ADV81155	Adv81155 Streptoco
38	91	13.1	470	8	ADT56185	Adt56185 Plant pol
39	91	13.1	484	3	AGG47777	Agg47777 Arabidops
40	90.5	13.0	639	9	ADM88474	Adm88474 Staphyloc
41	90	12.9	4544	8	ADP25443	Adp25443 Plasmodiu
42	89.5	12.9	292	8	ADQ66058	Adq66058 Novel hum
43	89.5	12.9	511	2	AAV35091	Aav35091 Chlamydia
44	89.5	12.9	645	9	ADW88441	Adw88441 Staphyloc
45	88.5	12.7	903	6	ABU24404	Abu24404 Protein e

#### ALIGNMENTS

#### RESULT 1

AAB48343  
ID AAB48343 standard; protein; 773 AA.

AC AAB48343;

XX 20-APR-2001 (first entry)

XX S. pneumoniae Spi30 polypeptide.

XX Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;  
XX bronchial; lung; blood; infection; immune response; immunotherapy;  
XX antibacterial; auditory; vaccine.

OS Streptococcus pneumoniae.

XX WO200076540-A2.

XX 21-DEC-2000.

XX 09-JUN-2000; 2000WO-US015925.

XX 10-JUN-1999; 99US-0138453P.

XX (MEDI-) MED IMMUNE INC.

PI Adamou JE, Choi GH;

XX WPI; 2001-112197/12.

DR N-PSDB; AAC84742.

XX New vaccines comprising Spi28 or Spi30 polypeptides, for treating and  
XX preventing pneumococcal infections, particularly infections caused by  
XX Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
XX blood infections.

XX Claim 8; Page 51-54; 54pp; English.

XX The invention relates to novel immunogenic polypeptides, Spi28 and Spi30  
XX from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
XX the treatment and prevention of pneumococcal infections, particularly  
XX infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
XX bronchial, lung or blood infections. The antigens are used as immunogenic  
XX agents to stimulate an immune response. The antisera and antibodies may  
XX also be used in diagnosing and treating pneumococcal infections.  
XX Recombinant polypeptides serve as a mechanism for stimulating production  
XX of antibodies for use in passive immunotherapy, diagnostic reagents, and

CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the S. pneumoniae Spl30 polypeptide  
XX  
SQ Sequence 773 AA;

Query Match 100.0%; Score 696; DB 4; Length 773;  
Best Local Similarity 100.0%; Pred. No. 1.1e-62;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 60  
DB 640 KMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 699  
QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDK 120  
DB 700 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDK 759  
QY 121 NNISKSSTNNPNK 134  
DB 760 NNISKSSTNNPNK 773

RESULT 2  
AAY81710  
ID AAY81710 standard; protein; 2120 AA.  
XX  
AC AAY81710;  
XX  
DT 02-JUN-2000 (first entry)  
XX  
DE Streptococcus pneumoniae protein sequence ID3.  
XX  
KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200006738-A2.  
XX  
PD 10-FEB-2000.  
XX  
PF 27-JUL-1999; 99WO-GB002452.  
XX  
PR 27-JUL-1999; 98GB-00016336.  
PR 19-MAR-1999; 99US-0125329P.  
XX  
PA (MICR-) MICROBIAL TECHNIQS LTD.  
XX  
PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
XX  
DR WPI; 2000-195301/17.  
DR N-PSDB; AA291806.  
XX  
PT Streptococcal proteins and polynucleotides useful for diagnosis,  
PT treatment and prophylaxis of bacterial infections.  
XX  
PS Claim 2; Page 41-42; 76pp; English.  
XX  
CC This sequence represents a Streptococcus pneumoniae protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
CC infection. As the sequences can be used to treat S. pneumoniae infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,

CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
CC meningitis  
XX  
SQ Sequence 2120 AA;

Query Match 100.0%; Score 696; DB 3; Length 2120;  
Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 60  
DB 1953 KMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 2012  
QY 61 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDK 120  
DB 2013 IEKKKEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDK 2072  
QY 121 NNISKSSTNNPNK 134  
DB 2073 NNISKSSTNNPNK 2086

RESULT 3  
ABU01020  
ID ABU01020 standard; protein; 2140 AA.  
XX  
AC ABU01020;  
XX  
DT 23-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
DE S. pneumoniae type 4 strain protein from coding region #590.  
XX  
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
KW gene therapy; vaccine.  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
PN WO200277021-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 27-MAR-2002; 2002WO-IB002163.  
XX  
PR 27-MAR-2001; 2001GB-00007658.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Masignani V, Tettelin H, Fraser C;  
XX  
DR WPI; 2003-040579/03.  
DR N-PSDB; ABX06302.  
XX  
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
PS Claim 1; SEQ ID NO 1180; 56pp; English.  
XX  
CC The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are the  
CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
CC sequence not defined in the specification, for amplifying a target

CC sequence contained within a Streptococcus nucleic acid sequence, where  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus  
 CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 XX  
 SQ Sequence 2140 AA;

Query Match 100.0%; Score 696; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGVYVNLKDTFIKPVFKK 60  
 DB 1973 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGVYVNLKDTFIKPVFKK 2032  
 QY 61 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDK 120  
 DB 2033 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDK 2092  
 QY 121 NNISKSSTTNNPNK 134  
 DB 2093 NNISKSSTTNNPNK 2106

RESULT 4  
 ABU45746  
 ID ABU45746 standard; protein; 2140 AA.

XX AC ABU45746;  
 XX DT 19-JUN-2003 (first entry)  
 XX DE Protein encoded by Prokaryotic essential gene #31273.  
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX OS Streptococcus pneumoniae.  
 XX FN WO20027183-A2.  
 XX PD 03-OCT-2002.  
 XX PF 21-MAR-2002; 2002WO-US009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 XX PR 06-SEP-2001; 2001US-00948993.  
 XX PR 25-OCT-2001; 2001US-0342923P.  
 XX PR 08-FEB-2002; 2002US-00072851.  
 XX PR 06-MAR-2002; 2002US-0362699P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI; 2003-029926/02.  
 XX DR N-PSDB; ACA49616.

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 73670; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 2140 AA;

Query Match 100.0%; Score 696; DB 6; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGVYVNLKDTFIKPVFKK 60  
 DB 1973 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKGAGVYVNLKDTFIKPVFKK 2032

QY 61 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDK 120  
 DB 2033 IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHHSQKSDTKDVTATVLDK 2092  
 QY 121 NNISKSSTTNNPNK 134  
 DB 2093 NNISKSSTTNNPNK 2106

RESULT 5  
 ADM92113  
 ID ADM92113 standard; protein; 2140 AA.

XX AC ADM92113;  
 XX DT 03-JUN-2004 (first entry)  
 XX DE S pneumoniae antigenic protein sequence SeqID310.  
 XX KW antibacterial; gene therapy; Streptococcus pneumoniae infection;  
 XX OS Streptococcus pneumoniae.

PN WO2004020609-A2.  
XX 11-MAR-2004.  
XX 02-SEP-2003; 2003WO-US027401.  
XX 30-AUG-2002; 2002US-0407082P.  
XX (TUFT ) UNIV TUFTS.  
XX Camilli A, Hava DL;  
PI WPI; 2004-239189/22.  
XX N-PSDB; ADM91876.  
XX New Streptococcus pneumoniae nucleic acid molecules, useful for  
PT diagnosing, treating and preventing active infections of Streptococcus  
PT pneumoniae.  
XX Claim 27; SEQ ID NO 310; 123pp; English.  
XX This invention relates to novel isolated Streptococcus pneumoniae nucleic  
CC acid molecules and the antigenic polypeptides encoded by them. The  
CC invention may be useful for the production of compounds with an  
CC antibacterial activity or for gene therapy. The nucleic acid molecules,  
CC compositions and methods disclosed are useful for treating Streptococcus  
CC pneumoniae infection. The present sequence is that of an S pneumoniae  
CC protein of the invention.  
XX  
SQ Sequence 2140 AA;  
Query Match 100.0%; Score 696; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKGDAGYVINLSKDTFIKPVFKK 60  
DB 1973 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKGDAGYVINLSKDTFIKPVFKK 2032  
QY 61 IEKKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120  
DB 2033 IEKKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2092  
QY 121 NNISKSTTNPNK 134  
DB 2093 NNISKSTTNPNK 2106  
RESULT 6  
ID ADT50099 standard; protein; 2140 AA.  
XX ADT50099;  
XX 13-JAN-2005 (first entry)  
XX S\_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.  
XX hyperimmune serum reactive antigen; antibacterial; vaccine;  
KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;  
KW sepsis; meningitis.  
XX Streptococcus pneumoniae TIGR4.  
XX WO2004092209-A2.  
PN 28-OCT-2004.  
XX 15-APR-2004; 2004WO-EP003984.  
XX 15-APR-2003; 2003EP-00450087.  
XX (INTE-) INTERCELL AG.

XX Meinke A, Nagy B, Hanner M, Dewasthaly S, Stierschneider U;  
XX WPI; 2004-758335/74.  
XX N-PSDB; ADT49955.  
XX New hyperimmune serum reactive antigens from Streptococcus pneumoniae,  
PT and encoding nucleic acid molecules, useful for diagnosing, preventing or  
PT treating S. pneumoniae infections.  
XX Disclosure; SEQ ID NO 177; 191pp; English.  
XX This invention relates to novel nucleic acids encoding hyperimmune serum  
CC reactive antigens, or fragments derived thereof. Specifically, it refers  
CC to antigens selected from peptides and serum reactive epitopes that can  
CC be used in pharmaceutical compositions that exhibit antibacterial  
CC activity. The present invention describes a composition (including the  
CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)  
CC that is useful for manufacturing a medicament such as a vaccine, which  
CC can be used to treat or prevent bacterial infections, particularly S.  
CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,  
CC bacteraemia sepsis and meningitis. The antigen or its fragment may also  
CC be used for isolating, purifying and/ or identifying an interaction  
CC partner of the hyperimmune serum reactive antigen, as well as for  
CC manufacturing a functional nucleic acid selected from aptamers and  
CC spiegelmers or for manufacturing a functional ribonucleic acid selected  
CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide  
CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen  
CC of the invention.  
XX  
SQ Sequence 2140 AA;  
Query Match 100.0%; Score 696; DB 8; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 4.5e-62;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKGDAGYVINLSKDTFIKPVFKK 60  
DB 1973 KEMSTTVSEEDFILPVYKGELEKGYQFDGWEISGFEGKGDAGYVINLSKDTFIKPVFKK 2032  
QY 61 IEKKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120  
DB 2033 IEKKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2092  
QY 121 NNISKSTTNPNK 134  
DB 2093 NNISKSTTNPNK 2106  
RESULT 7  
ID ADR94534 standard; protein; 637 AA.  
XX ADR94534;  
XX 16-DEC-2004 (first entry)  
XX Novel S. pneumoniae protein sequence, SEQ ID 3169.  
XX Meningitis; bacteraemia; pneumonia; otitis media; vaccine;  
KW bacterial infection.  
XX Streptococcus pneumoniae.  
XX US6800744-B1.  
XX 05-OCT-2004.  
XX 30-JUN-1998; 98US-00107433.  
XX 02-JUL-1997; 97US-0051553P.  
XX 12-MAY-1998; 98US-0085131P.  
XX



```
XX ADK48759;
AC
XX 20-MAY-2004 (first entry)
DT
XX Streptococcus pneumoniae protein, Seq ID No 5274.
DE
XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
KW
XX Streptococcus pneumoniae.
OS
XX US6699703-B1.
PN
XX 02-MAR-2004.
PD
XX 26-MAY-2000; 2000US-00583110.
PF
XX 02-JUL-1997; 97US-0051553P.
PR
XX 12-MAY-1998; 98US-0085131P.
PR
XX 30-JUN-1998; 98US-00107433.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
PI
XX WPI; 2004-212399/20.
DR
XX N-PSDB; ADK46098.
DR
XX
XX New nucleic acid molecules and polypeptides useful for diagnosing,
PT
XX preventing and treating pathological conditions resulting from bacterial
PT
XX infection, e.g. Streptococcus pneumoniae infection, and in drug
PT
XX screening.
XX
XX Disclosure; SEQ ID NO 5274; 301pp; English.
XX
XX The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC
XX and polypeptides. The nucleic acids and proteins are useful for
CC
XX diagnosing, preventing and treating pathological conditions resulting
CC
XX from bacterial infection, such as S. pneumoniae infection. These may also
CC
XX be used for drug screening procedures. The present sequence represents a
CC
XX Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC
XX data for this patent did not appear in the printed specification but was
CC
XX obtained in electronic format directly from USPTO at
CC
XX seqdata.uspto.gov/sequence.html.
XX
XX Sequence 2138 AA;
Query Match 99.6%; Score 693; DB 8; Length 2138;
Best Local Similarity 99.3%; Pred. No. 9.2e-62;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKK 60
Db 1771 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKK 2030
Qy 61 IEKKKEENKPTFDVSKKQPNVHNSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120
Db 2031 IEKKKEENKPTFDVSKKQPNVHNSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2090
Qy 121 NNISKSSTNNPNK 134
Db 2091 NNISKSSTNNPNK 2104
RESULT 10
AAW55096
ID AAW55096 standard; protein; 117 AA.
XX
XX AAW55096;
AC
XX 02-OCT-1998 (first entry)
DT
XX Streptococcus pneumoniae SP0043 protein.
DE
```

```
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX Streptococcus pneumoniae.
OS
XX WO9818930-A2.
PN
XX 07-MAY-1998.
PD
XX 30-OCT-1997; 97WO-US019422.
PF
XX 31-OCT-1996; 96US-0029960P.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
PI
XX WPI; 1998-272224/24.
DR
XX N-PSDB; AAV27357.
DR
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT
XX - or their epitope-containing fragments, useful in protective or
PT
XX therapeutic vaccines, and for diagnosis.
XX
XX Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
CC
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC
XX can be useful in vaccines for inducing protective antibodies against
CC
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC
XX are used to detect Streptococcus infection (by usual hybridisation or
CC
XX amplification methods), also for isolating Streptococcus genes or their
CC
XX allelic variants. The protein can be used similarly to detect specific
CC
XX antibodies in standard immunoassays, especially for diagnosing or
CC
XX monitoring infections. Antibodies which bind the protein are used to
CC
XX detect corresponding antigens, to purify the protein and for passive
CC
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
CC
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
CC
XX (especially 10-300) mu g/ml per dose
XX
XX Sequence 117 AA;
Query Match 88.4%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 2e-55;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 18 YKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 77
Db 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 60
Qy 78 KQNPQVNHNSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSSTNNPNK 134
Db 61 KQNPQVNHNSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKSSTNNPNK 117
RESULT 11
ABP54590
ID ABP54590 standard; protein; 117 AA.
XX
XX AC ABP54590;
AC
XX 04-SEP-2002 (first entry)
DT
XX
XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
DE
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
OS
XX US2002061545-A1.
PN
```

```

XX PD 23-MAY-2002.
XX PF
XX PP 22-JAN-2001; 2001US-00765272.
XX PR 30-OCT-1997; 97US-00961083.
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2002-479261/51.
XX DR N-PSDB; ABQ84625.
XX PT New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
XX PT and for preventing or attenuating disease caused by Streptococcus
XX PT infection.
XX PS Claim 11; Page 29; 70pp; English.
XX CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
XX CC pneumoniae antigens have antibacterial activity and can be used in
XX CC vaccines. The S. pneumoniae antigens can also be used to prevent or
XX CC attenuate a Streptococcal infection in an animal. The polynucleotides
XX CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
XX CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
XX CC of S. pneumoniae ORFs (open reading frames) which are used in an example
XX CC from the present invention
XX SQ Sequence 117 AA;
Query Match 88.4%; Score 615; DB 5; Length 117;
Best Local Similarity 100.0%; Pred. No. 2e-55;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 18 YKGELEKGYQPDGWEISGFEGKGDAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 77
Db 1 YKGELEKGYQPDGWEISGFEGKGDAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 60
Qy 78 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 134
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117
RESULT 12
ADC45149
ID ADC45149 standard; protein; 117 AA.
XX AC ADC45149;
XX DT 18-DEC-2003 (first entry)
XX DE S. pneumoniae antigenic protein SP043.
XX KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX OS Streptococcus pneumoniae.
XX FN US6573082-B1.
XX PD 03-JUN-2003.
XX PF 28-MAR-2000; 2000US-00536784.
XX PR 31-OCT-1996; 96US-0029960P.

```

```

PR 30-OCT-1997; 97US-00961083.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI; 2003-764574/72.
XX DR N-PSDB; ADC45148.
XX PT Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
XX PT useful for producing vaccines for prevention or attenuation of infection
XX PT by Streptococcus pneumoniae.
XX PS Example 1; SEQ ID NO 68; 58pp; English.
XX CC The invention relates to an isolated polynucleotide consisting of a
XX CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
XX CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
XX CC antigens. Also included are making a recombinant vector by inserting the
XX CC nucleic acid into a vector, an isolated polynucleotide consisting of at
XX CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
XX CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
XX CC acids are useful as DNA vaccine against Streptococcus pneumoniae
XX CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
XX CC antigen nucleic acids are useful as probes for use in diagnostic methods
XX CC for detecting S. pneumoniae gene expression. The present sequence
XX CC represents an S. pneumoniae antigenic protein.
XX SQ Sequence 117 AA;
Query Match 88.4%; Score 615; DB 7; Length 117;
Best Local Similarity 100.0%; Pred. No. 2e-55;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 18 YKGELEKGYQPDGWEISGFEGKGDAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 77
Db 1 YKGELEKGYQPDGWEISGFEGKGDAGYVINLSKDTFIKPVFKKIEEKKKEENKPTFDVSK 60
Qy 78 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 134
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117
RESULT 13
ADZ79639
ID ADZ79639 standard; protein; 188 AA.
XX AC ADZ79639;
XX DT 14-JUL-2005 (first entry)
XX DE P. falciparum merozoite surface protein 3, MSP3a to MSP3f fragment.
XX KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;
XX KW immunotherapy; malaria; antimalarial; vaccine.
XX OS Plasmodium falciparum.
XX FN WO2005040206-A1.
XX PD 06-MAY-2005.
XX PF 22-OCT-2004; 2004WO-EP012910.
XX PR 24-OCT-2003; 2003US-00691672.
XX PA (INSP ) INST PASTEUR.
XX PI Druilhe P;
XX DR WPI; 2005-355821/36.
XX

```

PT Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.

XX Disclosure: SEQ ID NO 7; 79pp; English.

XX The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID NO:1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID NO:2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also  
 CC described are: (i) a conjugate comprising the chimeric molecule of the  
 CC invention bound to a solid support, (ii) an immunogenic composition  
 CC comprising the chimeric molecule, the conjugate described above, or a  
 CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
 CC against malaria comprising the chimeric molecule, the conjugate described  
 CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
 CC association with a suitable vehicle, (iv) use of purified and/or  
 CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
 CC medicament against malaria, and (v) a medicament for passive  
 CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
 CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
 CC and MSP3 antigens are useful for the preparation of a vaccine composition  
 CC against malaria. This sequence represents Plasmodium falciparum MSP3a to  
 CC MSP3f fragment. Note: The present sequence given as SEQ ID NO:7 in the  
 CC Sequence Listing is not mentioned elsewhere in the specification.

XX Sequence 188 AA;

Query Match 15.4%; Score 107; DB 9; Length 188;  
 Best Local Similarity 23.8%; Pred. No. 0.012; 43; Indels 44; Gaps 6;  
 Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;  
 QY 1 KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKQDAG-----YVINLSKDTF 53  
 DB 19 KEASS-----YDYIL-----GWFFGGVPEHKKEENMLSHLYVSSKDKENI 59  
 QY 54 IKPVFKKIEEKKE-----ENKPTFDVSKKKNPQVNHSQLSHKSHKEDLQR 100  
 DB 60 SKENDVDLDEKEEAETEEELKEEETEESEISEDEEEEBEKEEBEENKKKQEK 119  
 QY 101 EHSOKSDSTKDVATVLDKNNISSKSTNN 131  
 DB 120 EQSNENNDDQKDMA-----QNLISGQNNN 145

RESULT 14  
 ADZ72253  
 ID ADZ72253 standard; protein; 354 AA.  
 XX  
 AC ADZ72253;  
 XX  
 DT 14-JUL-2005 (first entry)  
 XX  
 DE Plasmodium falciparum MSP-3-like protein, MSP-3-1 SEQ ID NO: 2.  
 XX  
 KW Nucleic acid vaccine; plasmodium falciparum infection; antimalarial;  
 KW infection; merozoite surface protein 3-like protein; MSP-3-1; antigen.  
 XX  
 OS Plasmodium falciparum.  
 XX  
 PN EP1526178-A1.  
 XX  
 PD 27-APR-2005.  
 XX  
 PF 24-OCT-2003; 2003EP-00292673.  
 XX  
 PR 24-OCT-2003; 2003EP-00292673.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX

PI Druilhe P;  
 XX WPI; 2005-323987/34.  
 DR N-PSDB; ADZ72252.

XX Novel MSP-3-like family genes located on chromosome 10 of Plasmodium  
 PT falciparum, which encode proteins useful for preparing vaccine  
 PT compositions against malaria.

XX Disclosure: SEQ ID NO 2; 137pp; English.

XX The present invention relates to the protection against malaria. More  
 CC particularly, the invention pertains to a family of MSP-3 (merozoite  
 CC surface protein 3)-like genes (MSP-3-1, MSP-3-2, MSP-3-3, MSP-3-4, MSP-3-  
 CC 5, MSP-3-6, MSP-3-7 and MSP-3-8) located on chromosome 10 of Plasmodium  
 CC falciparum, highly conserved in P. falciparum strains, simultaneously  
 CC expressed in P. falciparum at the erythrocytic stages and encoding  
 CC proteins which have a Asn-Leu-Arg-Asn or Asn-Leu-Arg-Lys signature at  
 CC their N-terminal extremity and which are located at the merozoite  
 CC surface. The characterization of this gene family enables the definition  
 CC of immunogenic and vaccine compositions against P. falciparum. The  
 CC present sequence is the P. falciparum MSP-3-1 protein.

XX Sequence 354 AA;

Query Match 15.4%; Score 107; DB 9; Length 354;  
 Best Local Similarity 23.8%; Pred. No. 0.03; 43; Indels 44; Gaps 6;  
 Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;  
 QY 1 KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKQDAG-----YVINLSKDTF 53  
 DB 185 KEASS-----YDYIL-----GWFFGGVPEHKKEENMLSHLYVSSKDKENI 225  
 QY 54 IKPVFKKIEEKKE-----ENKPTFDVSKKKNPQVNHSQLSHKSHKEDLQR 100  
 DB 226 SKENDVDLDEKEEAETEEELKEEETEESEISEDEEEEBEKEEBEENKKKQEK 285  
 QY 101 EHSOKSDSTKDVATVLDKNNISSKSTNN 131  
 DB 286 EQSNENNDDQKDMA-----QNLISGQNNN 311

RESULT 15  
 ADZ79634  
 ID ADZ79634 standard; protein; 169 AA.  
 XX  
 AC ADZ79634;  
 XX  
 DT 14-JUL-2005 (first entry)  
 XX  
 DE P. falciparum merozoite surface protein 3, amino acid residues 212-380.  
 XX  
 KW immune stimulation; fusion protein; merozoite surface protein 3; MSP3;  
 KW immunotherapy; malaria; antimalarial; vaccine.  
 XX  
 OS Plasmodium falciparum.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..169  
 FT /note= "Amino acid residues 212-380 of MSP3"  
 XX  
 PN WO2005040206-A1.  
 XX  
 PD 06-MAY-2005.  
 XX  
 PF 22-OCT-2004; 2004WO-EP012910.  
 XX  
 PR 24-OCT-2003; 2003US-00691672.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX  
 PI Druilhe P;  
 XX



DR WPI; 2005-355821/36.  
 XX Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.  
 XX  
 PS Claim 2; SEQ ID NO 2; 79pp; English.  
 XX  
 CC The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also  
 CC described are: (i) a conjugate comprising the chimeric molecule of the  
 CC invention bound to a solid support, (ii) an immunogenic composition  
 CC comprising the chimeric molecule, the conjugate described above, or a  
 CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
 CC against malaria comprising the chimeric molecule, the conjugate described  
 CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
 CC association with a suitable vehicle, (iv) use of purified and/or  
 CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
 CC medicament against malaria, and (v) a medicament for passive  
 CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
 CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
 CC and MSP3 antigens are useful for the preparation of a vaccine composition  
 CC against malaria. This sequence represents Plasmodium falciparum MSP3  
 CC protein (amino acid residues 212-380).  
 XX  
 SQ Sequence 169 AA;  
 Query Match 14.9%; Score 103.5; DB 9; Length 169;  
 Best Local Similarity 25.2%; Pred. No. 0.025;  
 Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;  
 Qy 1 KEMSTIVSEDFILPVYKGELEKGYQDFGWEISGF--EGKKDAG-----YVINLSKDTFP 53  
 Db 1 KEASS-----YDYIL-----GWFGGVPPEHKKENMLSHLYVSSKDKENI 41  
 Qy 54 IKPVFKIEKKKEENKPTFDVSKKKNPQVNHSQLN-----ESHRKEDIQR 100  
 Db 42 SKENDVDLDE--KKEEABETEEELKEEKEETETSEISEDEEEEEEKKEEKKKEOEK 100  
 Qy 101 EEHSQKSDSTKDVATVLDKNNISSKSTTN 131  
 Db 101 EQSNENNQDKDMEA-----QNLISKQNNN 126  
 RESULT 16  
 ADZ79635  
 ID ADZ79635 standard; protein; 647 AA.  
 XX  
 AC ADZ79635;  
 XX  
 DT 14-JUL-2005 (first entry)  
 XX  
 DE P. falciparum GLURP-MSP3 fusion protein.  
 XX  
 KW immune stimulation; fusion protein; glutamate-rich protein; GLURP;  
 KW merozoite surface protein 3; MSP3; immunotherapy; malaria; antimalarial;  
 KW vaccine.  
 XX  
 OS Plasmodium falciparum.  
 OS Synthetic.  
 XX  
 XX WO2005040206-A1.  
 PN  
 PD 06-MAY-2005.  
 XX  
 XX 22-OCT-2004; 2004WO-EP012910.  
 PF  
 XX 24-OCT-2003; 2003US-00691672.  
 PR

XX (INSP ) INST PASTEUR.  
 XX  
 PI Druilhe P;  
 XX  
 DR WPI; 2005-355821/36.  
 DR N-FSDB; ADZ79636.  
 XX  
 PT Chimeric molecule useful for preparing vaccine composition against  
 PT malaria, comprises glutamate-rich protein GLURP and Merozoite surface  
 PT protein 3 MSP3 moieties, and raises antibodies against moieties in mice  
 PT immunized with molecule.  
 XX  
 PS Disclosure; SEQ ID NO 3; 79pp; English.  
 XX  
 CC The invention relates to a chimeric molecule that comprises a glutamate-  
 CC rich protein (GLURP) moiety consisting of a polypeptide fragment (amino  
 CC acid residues 25-514) of GLURP (given as SEQ ID No:1) and a Merozoite  
 CC surface protein 3 (MSP3) moiety consisting of amino acid residues 212-380  
 CC of MSP3 (given as SEQ ID No:2), wherein the chimeric molecule raises  
 CC antibodies against both polypeptides in mice immunized with it. Also  
 CC described are: (i) a conjugate comprising the chimeric molecule of the  
 CC invention bound to a solid support, (ii) an immunogenic composition  
 CC comprising the chimeric molecule, the conjugate described above, or a  
 CC mixture of GLURP and MSP3 antigens as an immunogen, (iii) a vaccine  
 CC against malaria comprising the chimeric molecule, the conjugate described  
 CC above, or a mixture of GLURP and MSP3 antigens as an immunogen, in  
 CC association with a suitable vehicle, (iv) use of purified and/or  
 CC recombinant anti-MSP3 and anti-GLURP antibodies, for the preparation of a  
 CC medicament against malaria, and (v) a medicament for passive  
 CC immunotherapy of malaria, comprising the anti-MSP3 and anti-GLURP  
 CC antibodies. The chimeric molecule of the invention or a mixture of GLURP  
 CC and MSP3 antigens are useful for the preparation of a vaccine composition  
 CC against malaria. This sequence represents Plasmodium falciparum GLURP(27-  
 CC 500)-MSP3(212-380) fusion protein.  
 XX  
 SQ Sequence 647 AA;  
 Query Match 14.9%; Score 103.5; DB 9; Length 647;  
 Best Local Similarity 25.2%; Pred. No. 0.15;  
 Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;  
 Qy 1 KEMSTIVSEDFILPVYKGELEKGYQDFGWEISGF--EGKKDAG-----YVINLSKDTFP 53  
 Db 479 KEASS-----YDYIL-----GWFGGVPPEHKKENMLSHLYVSSKDKENI 519  
 Qy 54 IKPVFKIEKKKEENKPTFDVSKKKNPQVNHSQLN-----ESHRKEDIQR 100  
 Db 520 SKENDVDLDE--KKEEABETEEELKEEKEETETSEISEDEEEEEEKKEEKKKEOEK 578  
 Qy 101 EEHSQKSDSTKDVATVLDKNNISSKSTTN 131  
 Db 579 EQSNENNQDKDMEA-----QNLISKQNNN 604  
 RESULT 17  
 ADO19012  
 ID ADO19012 standard; protein; 651 AA.  
 XX  
 AC ADO19012;  
 XX  
 DT 12-AUG-2004 (first entry)  
 XX  
 DE Amino acid sequence for P. falciparum GLURP-MSP3 hybrid.  
 XX  
 KW Glutamate-rich protein; GLURP-MSP3 fusion protein;  
 KW merozoite surface protein 3; malarial vaccine; malaria; immune response;  
 KW antimalarial; immunostimulant.  
 XX  
 OS Plasmodium falciparum.  
 OS Synthetic.  
 XX  
 PN WO2004043488-A1.





PD 11-MAY-2000.

XX 05-NOV-1999; 99WO-US026796.

XX 05-NOV-1998; 98US-0107131P.

XX (HOFF/) HOFFMAN S.

PA (CARU/) CARUCCI D.

PA (GARD/) GARDNER M.

PA (VENT/) VENTER J C.

XX Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI; 2000-365347/31.

XX Proteins encoded by chromosome 2 of the human malarial parasite,

PT Plasmodium falciparum, useful as antimalarial vaccines and in the

PT diagnosis of P.falciparum infection.

XX Disclosure; Page 321-322; 577pp; English.

XX The present invention describes proteins and their fragments (I) encoded

CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.

CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)

CC vaccines against P. falciparum infection comprising (I) or (II). (I) and

CC (II) are useful for the development of vaccines against P. falciparum

CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to

CC immunogens comprising the sequences of (I), are useful in the detection

CC of infection with P. falciparum. Furthermore, (I) (especially when they

CC are refined or secreted or membrane proteins) can aid the identification

CC of drugs to treat or prevent P. falciparum infection, or they can be used

CC to identify drug resistance in P. falciparum. Sequencing of the

CC Plasmodium chromosome 2 and the subsequent identification of proteins

CC encoded by it will help to expand our understanding of parasite biology,

CC a process hampered by the complexity of the parasitic lifecycle, and

CC provide new targets for vaccine and drug development. Parasite resistance

CC to drugs and mosquito resistance to insecticides have led to a resurgence

CC of malaria in many parts of the world, and there is a pressing need for

CC vaccines and new drugs. AAA70078 and AAA70287 and AAB18144 to AAB18352

CC represent nucleotide and protein sequences given in the present

CC invention, but which are not specifically mentioned within the

CC specification

XX Sequence 665 AA;

XX Query Match 14.1%; Score 98; DB 3; Length 665;

XX Best Local Similarity 27.1%; Pred. No. 0.59;

XX Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;

QY 23 EKGYPDGMWEI--SGFEGKKDAGVYVNLTKDTFTKPVFKIEKKK-----BENK 70

DB 158 EKGKQ-----DISNSNAENKKD-----VKEGVKELEERKKBEKISDDHKVBEK 201

QY 71 PTFD----VSKKKONPQVNHSQLNESHKEDLQR--EHSQKSDSTKQVATATVLDKNNISS 125

DB 202 KSDDHKVBEENKSSDDHKVBEENKSSDDHKIEEVKVEEHEDEE-----DKKEKKS 252

QY 126 KSTTNPNK 134

DB 253 ENKNKDENK 261

RESULT 22

ABO23606

ID ABO23606 standard; protein; 665 AA.

XX ABO23606;

AC ABO23606;

DT 04-SEP-2003 (first entry)

XX Plasmodium falciparum outlier protein #3.

DE Candidate protein identification; pathogen; anti-infective;

XX

KW outlier protein; virulence protein; antigen; drug target protein;

XX pathogenic organism; antimicrobial.

XX Plasmodium falciparum.

OS US2003039963-A1.

XX 27-FEB-2003.

XX 30-MAR-2001; 2001US-00820843.

XX 30-MAR-2001; 2001US-00820843.

XX (BRAH/) BRAHMACHARI S K.

PA (RAMA/) RAMACHANDRAN S.

PA (NAND/) NANDI T.

PA (BHIM/) BHIMARAO C.

XX Brahmachari SK, Ramachandran S, Nandi T, Bhimmarao C;

XX WPI; 2003-492159/46.

XX Identifying candidate proteins useful as anti-infectives involves

PT matching outlier protein sequences with protein sequences in databases.

XX Example 7; Page 91-93; 117pp; English.

XX The present invention relates to a method for identifying candidate

CC proteins in pathogens useful as anti-infectives. The invention discloses

CC a computational method which involves the calculation of several sequence

CC attributes and their subsequent analysis results in the identification

CC of outlier proteins in different pathogens. The method is useful for the

CC identification of outlier proteins (e.g. virulence proteins, antigens or

CC proteins used as drug targets) in pathogenic organisms. The method of the

CC invention provides reproducible results as it does not depend on the

CC variable biochemical characterisation of proteins. ABO23500-ABO23617

CC represent outlier proteins identified from different pathogenic organisms

XX Sequence 665 AA;

XX Query Match 14.1%; Score 98; DB 7; Length 665;

XX Best Local Similarity 27.1%; Pred. No. 0.59;

XX Matches 35; Conservative 23; Mismatches 29; Indels 42; Gaps 7;

QY 23 EKGYPDGMWEI--SGFEGKKDAGVYVNLTKDTFTKPVFKIEKKK-----BENK 70

DB 158 EKGKQ-----DISNSNAENKKD-----VKEGVKELEERKKBEKISDDHKVBEK 201

QY 71 PTFD----VSKKKONPQVNHSQLNESHKEDLQR--EHSQKSDSTKQVATATVLDKNNISS 125

DB 202 KSDDHKVBEENKSSDDHKVBEENKSSDDHKIEEVKVEEHEDEE-----DKKEKKS 252

QY 126 KSTTNPNK 134

DB 253 ENKNKDENK 261

RESULT 23

ADP25441

ID ADP25441 standard; protein; 1791 AA.

XX ADP25441;

AC ADP25441;

XX 09-SEP-2004 (first entry)

DT Plasmodium falciparum antigen amino acid sequence SEQ ID NO:18.

XX Plasmodium falciparum; malaria parasite; antigen; malaria; immunogenic;

XX immune response; cytostatic; anti-HIV; virucide; hepatotropic;

XX antibacterial; vaccine; cancer; infectious disease; AIDS; hepatitis;

XX bacterial infection.

XX Plasmodium falciparum.

OS

XX WO2004053086-A2.  
 XX 24-JUN-2004.  
 XX 08-DEC-2003; 2003WO-US038966.  
 XX 06-DEC-2002; 2002US-0431494P.  
 XX (EPIM-) EPIMUNE INC.  
 XX (USNA ) US SEC OF NAVY.  
 XX Sette A, Doolan DL, Carucci DJ, Sidney J, Southwood S;  
 XX WPI; 2004-468856/44.  
 XX New isolated and/or purified Plasmodium falciparum polynucleotide  
 PT sequences, useful in inducing an immune response for preventing and/or  
 PT treating cancer and infectious diseases, such as AIDS, hepatitis, and  
 PT bacterial infections.  
 XX Claim 22; SEQ ID NO 18; 253pp; English.  
 XX The present invention describes an isolated and/or purified Plasmodium  
 CC falciparum (malaria parasite) antigen polynucleotide sequence, encoding  
 CC an immunogenic peptide. Also described: (1) a primer or detection probe  
 CC for hybridisation with a target sequence or the amplicon generated from a  
 CC target sequence comprising a sequence of at least 8-30, 35, 40, 45, 50,  
 CC 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 consecutive nucleotides of any  
 CC of the polynucleotide sequences as described above; (2) a DNA chip  
 CC comprising any of the polynucleotide sequences described above; (3) a  
 CC vector comprising a promoter operably linked to any of the nucleic acid  
 CC sequences described above; (4) a host cell transformed by the vector of  
 CC (3) or the polynucleotide described above; (5) a composition comprising a  
 CC carrier and the polynucleotide described above; (6) a method of inducing  
 CC an immune response in an individual comprising the administration of the  
 CC composition of (5) to induce an immune response; (7) an isolated  
 CC polypeptide comprising any of the amino acid sequences as encoded by the  
 CC polynucleotide described above; (8) a composition comprising a carrier  
 CC and the polypeptide of (7); (9) a method of detecting P. falciparum in  
 CC biological samples, comprising contacting a biological sample with the  
 CC isolated polynucleotide and detecting the hybridisation of the isolated  
 CC polynucleotides with nucleic acids contained in the sample; (10) a method  
 CC for eliciting an immune response in an individual, comprising the  
 CC administration of a composition comprising the polypeptides of (7) to an  
 CC individual to induce an immune response in the individual; (11) an  
 CC antibody that specifically binds to the P. falciparum polypeptide of (7);  
 CC and (12) detecting P. falciparum antigens, comprising contacting a sample  
 CC from a subject with the polypeptide of (7) and detecting the presence of  
 CC an antigen-antibody complex or detecting the stimulation of T-cells in  
 CC the sample. The P. falciparum antigens and immunogenic peptides have  
 CC cytotstatic, anti-HIV, virucide, hepatotropic and antibacterial  
 CC activities, and can be used in vaccines. The methods and compositions of  
 CC the present invention are useful for inducing an immune response for the  
 CC prevention and/or treatment of cancer and infectious diseases, such as  
 CC AIDS, hepatitis, and bacterial infections. The present sequence  
 CC represents a P. falciparum antigen amino acid sequence, which is used in  
 CC the exemplification of the present invention.  
 XX Sequence 1791 AA;  
 SQ

Query Match 13.4%; Score 93; DB 8; Length 1791;  
 Best Local Similarity 24.7%; Pred. No. 7.5;  
 Matches 45; Conservative 29; Mismatches 58; Indels 50; Gaps 7;  
 QY 1 KEMSTIVSEEDPILPVTKGELEKGYQDFGWEISG---FEGKDGAGYVNIISKDTFIPKP 57  
 DB 664 KBEIKTIVSDDMFTSPVNIKEYNNEQERKKEIVGNLSYDKTKIFPPIKFTKEGRKK- 722  
 QY 58 PKKIEEKEENK-----PTF-----DVSKKDNQV 84  
 DB 723 -KKIEKKEKKEKNNNNFLYNDYSYSSPKYGDNNNNFVIYIRKDKPQKPDHPNF 781

QY 85 NHSQI-----NESHK-----EDLQREHSHQSDSTKD-VTATVLKNNISKSSTNNP 132  
 DB 782 NFSKFLHYNPMKKNKKNKKNKKNRNEYNTTSSKDGYSYNFLSDLSFSSDNEYSSD 841  
 QY 133 NK 134  
 DB 842 NE 843  
 RESULT 24  
 ABR64281  
 ID ABR64281 standard; protein; 2468 AA.  
 XX AC ABR64281;  
 XX DT 15-OCT-2003 (first entry)  
 XX DE Angiogenesis protein BNO382.  
 XX KW Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;  
 KW antipsoriatic; antiarteriosclerotic; cardiast; vasotropic; angiogenesis;  
 KW gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;  
 KW diabetic retinopathy; cardiovascular disease; atherosclerosis;  
 KW ischemic limb disease; coronary artery disease.  
 XX OS Homo sapiens.  
 XX WO2003027285-A1.  
 XX 03-APR-2003.  
 XX 19-SEP-2002; 2002WO-AU001282.  
 XX 27-SEP-2001; 2001AU-00007973.  
 XX 27-SEP-2001; 2001AU-00007974.  
 XX 11-OCT-2001; 2001AU-00008210.  
 XX 29-OCT-2001; 2001AU-00008532.  
 XX 13-NOV-2001; 2001AU-00008838.  
 XX 28-AUG-2002; 2002AU-00951032.  
 XX (BION-) BIONOMICS LTD.  
 XX Gamble JR, Hahn CN, Vadas MA;  
 XX WPI; 2003-354655/33.  
 XX N-PSDB; ACP34559.  
 XX New angiogenic genes and polypeptides, useful for diagnosing,  
 PT prognosticating or treating an angiogenesis-related disorder, e.g.  
 PT cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or  
 PT cardiovascular diseases.  
 XX Claim 15; SEQ ID NO 216; 90pp; English.  
 XX The invention relates to the isolation of novel genes (ACP34446-ACP34559)  
 CC encoding proteins (ABR64180-ABR64281) involved in the process of  
 CC angiogenesis. The nucleic acid molecules are useful in identifying and/or  
 CC obtaining full-length human genes involved in an angiogenic process. The  
 CC nucleic acid molecule, polypeptides or complexes encoded, calls or  
 CC genetically modified non-human animals derived from these are useful for  
 CC the screening of candidate pharmaceutical compounds used in treating  
 CC angiogenesis-related disorders. They are also useful for diagnosing,  
 CC prognosticating or treating an angiogenesis-related disorder, which  
 CC involves uncontrolled or enhanced angiogenesis or is a disorder in which  
 CC a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,  
 CC diabetic retinopathy, psoriasis or cardiovascular diseases such as  
 CC atherosclerosis), or involves inappropriately arrested or decreased  
 CC angiogenesis or is a disorder in which an expanding vasculature is of  
 CC benefit (e.g. ischemic limb disease or coronary artery disease). The  
 CC modulator of expression or activity of the polypeptide encoded by the  
 CC nucleic acid sequence is useful for manufacturing a medicament for the  
 CC treatment of an angiogenesis-related disorder. This sequence corresponds  
 CC to one of the novel angiogenic protein

```

XX SQ Sequence 2468 AA;
Query Match 13.3%; Score 92.5; DB 6; Length 2468;
Best Local Similarity 31.0%; Pred. No.13;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 48 LSKDTFIKPVFKTIEEKKEENKFTFDVSKKONPQVNHSQLNESHKEDLQRE-----E 102
Db 638 VKEETKVKP-----EDKKEEKEKPKKEVAKKEDKTPI---KKEKPKKEEVKKEVKKEIK 689
Qy 103 HSQKSDSTKDV 113
Db 690 KEEKKEPKKEV 700

RESULT 25
ADE62723
ID ADE62723 standard; protein; 2468 AA.
XX AC ADE62723;
XX DT 29-JAN-2004 (first entry)
XX Human Protein NP_005900, SEQ ID NO 8656.
XX Human; pain; neuronal tissue; gene therapy;
KW spinal segmental nerve injury; chronic constriction injury; CCI;
KW spared nerve injury; SNI; Chung.
XX OS Homo sapiens.
XX WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX (GEO ) GEN HOSPITAL CORP.
PA (PAB ) BAYER AG.
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WFI; 2003-268312/26.
XX GENBANK; NP_005900.
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX Claim 1; Page; 1017pp; English.
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more

```

CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
XX  
XX

SQ Sequence 2468 AA;

Query Match 13.3%; Score 92.5; DB 7; Length 2468;  
Best Local Similarity 31.0%; Pred.No. 13;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 48 LSKDTFIKPVFKIKIEKKKEENKPTFDVSKKKNPQVNHSQLNESHKRKDLORE-----E 102  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 638 VVKETKVKP-----EDKCEKPKPKKEVAKDKTPI---KKEEKPKGEVKKVEIK 689

Qy 103 HSOKSDSTKDV 113  
: : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 690 KEEKGPKKEV 700

RESULT 26  
ADE62719  
ID ADE62719 standard; protein; 2468 AA.  
XX  
AC ADE62719;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human Protein AAA18904, SEQ ID NO 8652.  
XX  
KW Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Homo sapiens.  
XX  
XX WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI; 2003-268312/26.  
DR GENBANK; AAA18904.  
XX  
FT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX

The invention discloses a composition comprising two or more isolated rat  
or human polynucleotides or a polynucleotide which represents a fragment,  
derivative or allelic variation of the nucleic acid sequence. Also  
claimed are a vector comprising the novel polynucleotide, a host cell  
comprising the vector, a method for identifying a nucleotide sequence  
which is differentially regulated in an animal subjected to pain and a  
kit to perform the method, an array, a method for identifying an agent  
that increases or decreases the expression of the polynucleotide sequence  
that is differentially expressed in neuronal tissue of a first animal



XX WIPI: 2003-268312/26.  
DR GENBANK; NP\_005900.  
XX  
XX New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human protein (shown in table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 2468 AA;

Query Match 13.3%; Score 92.5; DB 7; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 13;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
Qy 48 LSKDTFIKVPFKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQRE-----E 102  
Db 638 VKKETVKVP-----EDKKEEKPKKEVAKEDKTPI---KKEKPKKEVKKEVKKEIK 689  
Qy 103 HSQKSDSTKDV 113  
Db 690 KEEKKEPKKEV 700

RESULT 29  
ADL12997  
ID ADL12997 standard; protein; 2468 AA.

AC ADL12997;  
XX  
XX  
DT 06-MAY-2004 (first entry)  
XX  
XX Human steroid-induced C3A liver cell protein #112.  
DE  
XX  
XX Hepatotropic; Gene therapy; Wilson disease; liver disorder;  
KW steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.  
XX

OS Homo sapiens.  
XX  
XX US6673549-B1.  
PN  
XX  
XX 06-JAN-2004.  
PD  
XX  
XX 12-OCT-2001; 2001US-00976594.  
PF  
XX  
XX

PR 12-OCT-2000; 2000US-0240409P.  
XX  
XX (INCY-) INCYTE CORP.  
XX  
XX Furness LM, Buchbinder JL;  
PI  
XX  
XX WIPI; 2004-068610/07.  
DR  
XX  
XX

XX Combination useful for preparing a composition for treating liver  
PT disorders associated with steroid therapy, e.g., cirrhosis or hepatitis,  
PT comprises cDNAs that are differentially expressed in response to steroid  
PT treatment.  
XX  
XX Disclosure; SEQ ID NO 726; 141pp; English.

XX The invention relates to a combination comprising cDNAs that are  
CC differentially expressed in response to steroid treatment. Also included  
CC are the following: a high throughput method for using a cDNA to detect  
CC differential expression of nucleic acids in a sample; and a high  
CC throughput method of screening molecules or compounds to identify a  
CC ligand that specifically binds a cDNA. The sample is from a subject with  
CC Wilson disease and comparison of a standard defines a stage of that  
CC disease. The high throughput method of screening molecules or compounds  
CC to identify a ligand that specifically binds a cDNA comprises: combining  
CC the combination with molecules or compounds under conditions to allow  
CC specific binding; and detecting specific binding between each cDNA and at  
CC least one molecule or compound. The molecules or compounds are regulatory  
CC proteins. The combination is useful for preparing a composition for  
CC treating liver disorders associated with steroid therapy, e.g., cirrhosis  
CC or hepatitis. The present sequence represents a human protein which is  
CC differentially expressed in steroid-induced C3A liver cells. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html.

XX Sequence 2468 AA;

Query Match 13.3%; Score 92.5; DB 8; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 13;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
Qy 48 LSKDTFIKVPFKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQRE-----E 102  
Db 638 VKKETVKVP-----EDKKEEKPKKEVAKEDKTPI---KKEKPKKEVKKEVKKEIK 689  
Qy 103 HSQKSDSTKDV 113  
Db 690 KEEKKEPKKEV 700

RESULT 30  
ADN05260  
ID ADN05260 standard; protein; 2468 AA.

AC ADN05260;  
XX  
XX  
DT 01-JUL-2004 (first entry)  
XX  
XX Antipsoriatic protein sequence #805.  
DE  
XX  
XX antipsoriatic; gene therapy; psoriasis; diagnosis.

OS Homo sapiens.  
XX  
XX WO2004028479-A2.  
PN  
XX  
XX 08-APR-2004.  
PD  
XX  
XX 25-SEP-2003; 2003WO-US030907.  
PF  
XX  
XX 25-SEP-2002; 2002US-0414006P.  
PR  
XX  
XX (GETH ) GENENTECH INC.





```
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR N-PSDB; AAS80823.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 46995; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (II) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2519 AA;
SQ
Query Match 13.3%; Score 92.5; DB 4; Length 2519;
Best Local Similarity 31.0%; Pred. No. 13;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;
Oy 48 LSKDTFIKVPFKKIEKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQRE-----E 102
Db 689 VKKETKVKP-----EDKKEEKPKKEVAKKEDKTPI---KKEEKPKKEEVKKEVKEIK 740
Oy 103 HSQKSDSTKDV 113
Db 741 KEEKKEPKKEV 751
RESULT 33
ADN04561
ID ADN04561 standard; protein; 2527 AA.
XX
XX ADN04561;
XX
XX 01-JUL-2004 (first entry)
DT
XX Antipsoriatic protein sequence #471.
DE
XX antipsoriatic; gene therapy; psoriasis; diagnosis.
XX
XX Homo sapiens.
XX
XX WO2004028479-A2.
PN
Query Match 13.3%; Score 92.5; DB 8; Length 2527;
Best Local Similarity 31.0%; Pred. No. 13;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;
Oy 48 LSKDTFIKVPFKKIEKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQRE-----E 102
Db 697 VKKETKVKP-----EDKKEEKPKKEVAKKEDKTPI---KKEEKPKKEEVKKEVKEIK 748
Oy 103 HSQKSDSTKDV 113
Db 749 KEEKKEPKKEV 759
RESULT 34
ADS93954
ID ADS93954 standard; protein; 635 AA.
XX
XX ADS93954;
XX
XX 02-DEC-2004 (first entry)
DT
XX Fibrinogen-binding polypeptide, SEQ ID No 19.
XX
XX Fibrinogen-binding; adhesion factor; vaccine; bacterial infection;
XX Streptococcus agalactiae infection; antibacterial; gene therapy;
XX ribozyme; antisense; siRNA; anticaline; aptamer; epiegelmer.
XX
XX Streptococcus agalactiae.
XX
XX WO2004035618-A2.
XX
XX 29-APR-2004.
XX
XX 15-OCT-2003; 2003WO-EP011436.
XX
XX 15-OCT-2002; 2002EP-00023141.
PR 20-MAR-2003; 2003EP-00006393.
XX
XX (INTE-) INTERCELL AG.
XX
XX Reinscheid DJ, Gutekunst H, Schubert A, Eikmanns BJ, Meinke A;
XX
XX WPI; 2004-357201/33.
XX
```

PT New nucleic acid molecules and encoded adhesion factors and/or fibrinogen  
 PT -binding polypeptides for diagnosing, preventing or treating bacterial  
 PT infections, preferably Streptococcus agalactiae infection.

PS Claim 13; SEQ ID NO 19; 225pp; English.

XX The invention relates to a novel isolated nucleic acid molecule encoding  
 CC a fibrinogen-binding polypeptide or its fragment, or an adhesion factor  
 CC or its fragment. The invention further comprises: a vector comprising the  
 CC above nucleic acid molecule; a cell, preferably a host cell, comprising  
 CC the vector; a polypeptide, preferably a fibrinogen-binding polypeptide  
 CC and/or an adhesion factor, comprising an amino acid sequence encoded by  
 CC the above nucleic acid molecule; a process for producing the above  
 CC polypeptide or its fragment; a process for producing a cell that  
 CC expresses the polypeptide or its fragment; a pharmaceutical composition,  
 CC especially a vaccine, comprising the polypeptide or its fragment, or the  
 CC above nucleic acid molecule; an antibody, or its part, that binds to at  
 CC least a selective part of the polypeptide or its fragment; methods for  
 CC identifying an antagonist capable of reducing or inhibiting the activity  
 CC of the polypeptide or its fragment, capable of binding to the  
 CC polypeptide, or capable of reducing or inhibiting the interaction  
 CC activity of the polypeptide or its fragment to its interaction partner;  
 CC an antagonist identified by the above method; processes for in vitro  
 CC diagnosis of a bacterial infection, preferably Streptococcus agalactiae  
 CC infection, or a disease related to the expression of the above  
 CC polypeptide or its fragment; and an affinity device comprising a support  
 CC material and immobilized to the support material the above polypeptide or  
 CC nucleic acid molecule. The fibrinogen-binding polypeptide has  
 CC antibacterial activity. The nucleic acid may be used in gene therapy to  
 CC treat disorders. The nucleic acid is used for manufacturing or generating  
 CC functional ribonucleic acids selected from ribozymes, antisense nucleic  
 CC acids and siRNA. The polypeptide or its fragment is useful for  
 CC manufacturing a medicament, especially a vaccine against bacterial  
 CC infection. The polypeptide is also used for manufacturing an antibody  
 CC that may be used for manufacturing a medicament for treating or  
 CC preventing bacterial infections, especially S. agalactiae infections. The  
 CC polypeptide is also used as an antigen, or for isolating, purifying  
 CC and/or identifying an interaction partner of the polypeptide or its  
 CC fragment, or for generating a peptide binding to the polypeptide, such as  
 CC anticatalines. It may also be used for manufacturing or generating a  
 CC functional nucleic acid selected from aptamers and spiegelmers. This  
 CC sequence represents a fibrinogen-binding polypeptide of the invention.

XX Sequence 635 AA;

Query Match 13.1%; Score 91.5; DB 8; Length 635;  
 Best Local Similarity 24.5%; Pred. No. 2.6;  
 Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 7;  
 Qy 3 MSSTIVSEDFILP--VYKGEI-----EKGYQFD-----GWEISGFEGKKDAGYVINLSK 50  
 Db 317 LTSYLENKEKFLVPNIPYKNKILREEDKYSFEDDEBEFGNELLSYNKLKNEVLVPVNITT 376  
 Qy 51 DTFIKPVFKKIEEKKERENKPTFVSKKDNQVNHSQLNESHKEDLQREHS--QKSD 108  
 Db 377 STILKP-----FEQKIVED---FNPYSNLDNLEIKKIRLNGSQKQVEQKTSPTPOKE 429  
 Qy 109 STKDVATVLDKN--NISKSTT 129  
 Db 430 TVKEQTEQKVSNGTQVEKKSET 452

RESULT 35  
 ADV83292

ID ADV83292 standard; protein; 635 AA.

XX AC ADV83292;

XX DT 24-FEB-2005 (first entry)

XX DE Streptococcus agalactiae protein, SEQ ID 4433.

XX KW Antibacterial; vaccine; bacterial infection.

XX Streptococcus agalactiae.  
 XX WO200292818-A2.  
 XX PD 21-NOV-2002.

XX 26-APR-2002; 2002MO-IB003059.

XX 26-APR-2001; 2001FR-00005642.

XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalloui L;  
 XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 XX agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 XX and identification of therapeutic targets.

XX Claim 6; SEQ ID NO 4433; 439pp; French.

XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.

XX Sequence 635 AA;

Query Match 13.1%; Score 91.5; DB 8; Length 635;  
 Best Local Similarity 24.5%; Pred. No. 2.6;  
 Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 7;  
 Qy 3 MSSTIVSEDFILP--VYKGEI-----EKGYQFD-----GWEISGFEGKKDAGYVINLSK 50  
 Db 317 LTSYLENKEKFLVPNIPYKNKILREEDKYSFEDDEBEFGNELLSYNKLKNEVLVPVNITT 376  
 Qy 51 DTFIKPVFKKIEEKKERENKPTFVSKKDNQVNHSQLNESHKEDLQREHS--QKSD 108  
 Db 377 STILKP-----FEQKIVED---FNPYSNLDNLEIKKIRLNGSQKQVEQKTSPTPOKE 429  
 Qy 109 STKDVATVLDKN--NISKSTT 129  
 Db 430 TVKEQTEQKVSNGTQVEKKSET 452

RESULT 36  
 ADV89902

ID ADV89902 standard; protein; 643 AA.

XX AC ADV89902;

XX DT 24-FEB-2005 (first entry)

XX DE Streptococcus agalactiae protein sequence, SEQ ID 2296.

XX KW Antibacterial; Vaccine; bacterial infection.

```

XX Streptococcus agalactiae.
OS
XX
XX FR2824074-A1.
XX
XX 31-OCT-2002.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX (INSP ) INST PASTEUR.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Glaser P, Rusniok C, Chevalier P, Frangeul L, Lalioui L;
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.
XX
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips
XX and identification of therapeutic targets.
XX
XX Claim 6; SEQ ID NO 2296; 2687pp; French.
XX
XX The present invention relates to novel Streptococcus agalactiae
XX nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;
XX ADV87746-ADV89550). The nucleotide sequences encode polypeptides of S.
XX agalactiae involved in the synthesis of amino acids, cell membranes,
XX intermediate (central) metabolism, energetic metabolism, fatty acid and
XX phospholipid metabolism, nucleotide metabolism including purines,
XX pyrimidines and/or nucleosides, regulatory functions, replication,
XX transcription, translation, protein transport, adaptation to atypical
XX conditions, sensitivity to medicines and/or analogues, functions related
XX to transporters, biosynthesis of cofactors, prosthetic groups and
XX transporters, cell membrane proteins and cellular machinery. (I) are
XX useful for the detection and/or amplification of nucleic acids.
XX Pharmaceutical composition comprising (I) or (II) are useful for
XX treatment of a bacterial S. agalactiae infection. Note: WO20022818A2 is
XX equivalent for the present basic patent FR2824074A1. WO20022818A2
XX contains 6617 sequence whereas the present patent only contains 2344
XX sequences.
XX
XX Sequence 643 AA;
XX
Query Match 13.1%; Score 91.5; DB 8; Length 643;
Best Local Similarity 24.5%; Pred. No. 2.7;
Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 7;
QY 3 MSSTVSEEDFILP--VYKGEI-----EKGYQFD-----GWEISGFECKDAGYVINLSK 50
DB 325 LTSYLENKEKFLVNPINPKLLIREEDKYSEFDEDEFGNLLSYNKLKNEVLVPIVNTT 384
QY 51 DTFIKPVFKIEEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHS--QKSD 108
DB 385 STILKP----PEQKKIVED---FNPYSNLDNLEIKIRLNGSQKQKVEKTSPTPQKE 437
QY 109 STKQVATVLDKN--NISSKSTT 129
DB 438 TVKEQTEQKVSNGTQVEKKSET 460
RESULT 37
ADV81155
ID ADV81155 standard; protein; 643 AA.
XX
XX ADV81155;
XX
XX 24-FEB-2005 (first entry)
XX
XX Streptococcus agalactiae protein, SEQ ID 2296.
XX
XX Antibacterial; vaccine; bacterial infection.

```

---

```

XX Streptococcus agalactiae.
OS
XX
XX WO200292818-A2.
XX
XX 21-NOV-2002.
XX
XX 26-APR-2002; 2002WO-IB003059.
XX
XX 26-APR-2001; 2001FR-00005642.
XX
XX (INSP ) INST PASTEUR.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Glaser P, Rusniok C, Chevalier P, Frangeul L, Lalioui L;
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;
XX WPI; 2004-101891/11.
XX
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips
XX and identification of therapeutic targets.
XX
XX Claim 6; SEQ ID NO 2296; 439pp; French.
XX
XX The present invention relates to novel Streptococcus agalactiae
XX nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and
XX novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The
XX nucleotide sequences encode polypeptides of S. agalactiae involved in the
XX synthesis of amino acids, cell membranes, intermediate (central)
XX metabolism, energetic metabolism, fatty acid and phospholipid metabolism,
XX nucleotide metabolism including purines, pyrimidines and/or nucleosides,
XX regulatory functions, replication, transcription, translation, protein
XX transport, adaptation to atypical conditions, sensitivity to medicines
XX and/or analogues, functions related to transporters, biosynthesis of
XX cofactors, prosthetic groups and transporters, cell membrane proteins and
XX cellular machinery. (I) are useful for the detection and/or amplification
XX of nucleic acids. Pharmaceutical composition comprising (I) or (II) are
XX useful for treatment of a bacterial S. agalactiae infection. The complete
XX genome of Streptococcus agalactiae is given in ADV81204. Note: The
XX present patent is an equivalent for the basic patent FR2824074A1, which
XX contains only 2344 sequences.
XX
XX Sequence 643 AA;
XX
Query Match 13.1%; Score 91.5; DB 8; Length 643;
Best Local Similarity 24.5%; Pred. No. 2.7;
Matches 35; Conservative 32; Mismatches 53; Indels 23; Gaps 7;
QY 3 MSSTVSEEDFILP--VYKGEI-----EKGYQFD-----GWEISGFECKDAGYVINLSK 50
DB 325 LTSYLENKEKFLVNPINPKLLIREEDKYSEFDEDEFGNLLSYNKLKNEVLVPIVNTT 384
QY 51 DTFIKPVFKIEEKKKEENKPTFDVSKKONPQVNHSQLNESHKREDLQREHS--QKSD 108
DB 385 STILKP----PEQKKIVED---FNPYSNLDNLEIKIRLNGSQKQKVEKTSPTPQKE 437
QY 109 STKQVATVLDKN--NISSKSTT 129
DB 438 TVKEQTEQKVSNGTQVEKKSET 460
RESULT 38
ADT56185
ID ADT56185 standard; protein; 470 AA.
XX
XX ADT56185;
XX
XX 13-JAN-2005 (first entry)
XX
XX Plant polypeptide, SEQ ID 6262.
XX
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;

```

KW disease resistance; galactomannan production; plant growth regulator;  
KW heat tolerance; herbicide tolerance; lignin production;  
KW extreme osmotic condition tolerance; pathogens resistance;  
KW pest resistance; yield improvement; seed oil yield; seed protein yield.  
XX Viridiplantae.  
OS

XX US2004216190-A1.

XX 28-OCT-2004.

XX 18-DEC-2003; 2003US-00739930.

XX 28-APR-2003; 2003US-00424599.

XX 28-APR-2003; 2003US-00425115.

XX (KOVA/) KOVALIC D K.

XX Kovalic DK;

XX WPI; 2004-757369/74.

XX New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.

XX Claim 2; SEQ ID NO 6262; 14pp; English.

XX The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5544 nucleotide sequences (CDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The CDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and rape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant protein  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.

XX Sequence 470 AA;

Query Match 13.1%; Score 91; DB 8; Length 470;

Best Local Similarity 19.3%; Score No. 2;

Matches 29; Conservative 29; Mismatches 58; Indels 34; Gaps 4;

Oy 18 YKGELEKGYQFDGWISGFE-----GKKDAGYVNLKSDTFIKPVFKIKBEKEEN 69

Db 100 YVQDLARRIYDE-EATGSSQAQRIDHPNCKNVGITKAFENSPIEETSHRVDNDKQINN 158  
Qy 70 KPTFDVSKKDN-----PQVNHSQLNE-----SHRKEDLQREHS 104  
Db 159 QKNFTAAKSSSENAVSRVSPGADHKRAEVMGKPMENRDQVQTESAESKSHRKENVIKSEK 218  
Qy 105 QKSDSTKDVTTATVLDKNNISKSTNNPNK 134  
Db 219 RDQGVKTKAKDKRKNKEKKEKTESINK 248

RESULT 39

AAG47777

ID AAG47777 standard; protein; 484 AA.

XX AAG47777;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 60255.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 09-MAR-1999; 99US-0123180P.

XX 23-MAR-1999; 99US-0123548P.

XX 25-MAR-1999; 99US-0125788P.

XX 29-MAR-1999; 99US-0126264P.

XX 01-APR-1999; 99US-0126785P.

XX 06-APR-1999; 99US-0127462P.

XX 08-APR-1999; 99US-0128234P.

XX 16-APR-1999; 99US-0128714P.

XX 19-APR-1999; 99US-0129845P.

XX 21-APR-1999; 99US-0130077P.

XX 23-APR-1999; 99US-0130510P.

XX 28-APR-1999; 99US-0130851P.

XX 30-APR-1999; 99US-0131449P.

XX 04-MAY-1999; 99US-0132048P.

XX 05-MAY-1999; 99US-0132484P.

XX 06-MAY-1999; 99US-0132485P.

XX 07-MAY-1999; 99US-0132487P.

XX 11-MAY-1999; 99US-0132863P.

XX 14-MAY-1999; 99US-0134256P.

XX 14-MAY-1999; 99US-0134218P.

XX 14-MAY-1999; 99US-0134219P.

XX 14-MAY-1999; 99US-0134221P.

XX 18-MAY-1999; 99US-0134370P.

XX 19-MAY-1999; 99US-0134768P.

XX 20-MAY-1999; 99US-0134941P.

XX 21-MAY-1999; 99US-0135124P.

XX 24-MAY-1999; 99US-0135353P.

XX 25-MAY-1999; 99US-0135629P.

XX 27-MAY-1999; 99US-0136021P.

XX 27-MAY-1999; 99US-0136392P.

XX 01-JUN-1999; 99US-0136782P.

XX 03-JUN-1999; 99US-0137222P.

XX 04-JUN-1999; 99US-0137528P.

XX 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139111P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139898P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0145986P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.

PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161040P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 13.1%; Score 91; DB 3; Length 484;  
Best Local Similarity 19.3%; Pred. No. 2;  
Matches 29; Conservative 29; Mismatches 58; Indels 34; Gaps 4;



**THIS PAGE BLANK (user)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:41:16 ; Search time 87.4631 Seconds  
(without alignments)  
1161.588 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTTTQNGKMSSTIVSE.....ATVLDKNNISSKSTNNPNK 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt.05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	748	100.0	2119	2	Q9AHT5 STRPN	Q9Aht5 streptococc
2	748	100.0	2140	2	Q97RY6 STRPN	Q97ry6 streptococc
3	745	99.6	2144	2	Q8DQP7 STRR6	Q8dqp7 streptococc
4	744	99.5	2144	2	Q9S4M8 STRPN	Q9s4m8 streptococc
5	128.5	17.2	300	2	Q4XUI6 PLACH	Q4xui6 plasmodium
6	112	15.0	361	2	Q95PI5 PLASMODIUM	Q95pi5 plasmodium
7	110	14.7	346	2	Q9U0G0 PLARE	Q9u0g0 plasmodium
8	110	14.7	379	2	Q25705 PLAPA	Q25705 plasmodium
9	110	14.7	775	2	Q8CPK8 STAP	Q8cpk8 staphylococ
10	109	14.6	3008	2	Q81436 PLAPF7	Q81436 plasmodium
11	108.5	14.5	600	2	Q77355 PLAPF7	Q77355 plasmodium
12	108.5	14.5	1038	2	Q90784 CHICK	Q90784 gallus gall
13	108	14.4	354	2	Q25995 PLAPF7	Q25995 plasmodium
14	108	14.4	354	2	Q81J55 PLAPF7	Q81j55 plasmodium
15	108	14.4	384	2	Q50VJ0 ENTHI	Q50vj0 entamoeba h
16	107	14.3	829	2	Q815F3 PLAPF7	Q815f3 plasmodium
17	106.5	14.2	374	2	Q5V9M0 PLAKN	Q5v9m0 plasmodium
18	106.5	14.2	470	2	Q9FJK9 ARATH	Q9fjk9 arabidopsis
19	106	14.2	616	2	Q6BRW2 DEBHA	Q6brw2 debaromyce
20	104.5	14.0	379	2	Q9U6C4 PLAPA	Q9u6c4 plasmodium
21	104.5	14.0	380	2	Q26019 PLAPA	Q26019 plasmodium
22	104.5	14.0	736	2	Q4VYV2 PLABE	Q4vyv2 plasmodium
23	104	13.9	662	2	Q4YMU4 PLABE	Q4ymu4 plasmodium
24	104	13.9	951	2	Q96228 PLAPF7	Q96228 plasmodium
25	103.5	13.8	325	2	Q44016 DICDI	Q44016 dictyosteli
26	103.5	13.8	379	2	Q25706 PLAPA	Q25706 plasmodium
27	103	13.8	296	2	Q50LX8 ENTHI	Q50lx8 entamoeba h
28	102.5	13.7	775	2	Q5HQ11 STAEQ	Q5hq11 staphylococ
29	102.5	13.7	1069	2	Q512T7 ENTHI	Q512t7 entamoeba h
30	102.5	13.7	2081	2	Q9LH98 ARATH	Q9lh98 arabidopsis
31	102	13.6	540	2	Q54MT2 DICDI	Q54mt2 dictyosteli

32	101.5	13.6	382	2	Q9V7J0 DROME	Q9v7j0 drosophila
33	101.5	13.6	500	2	Q6BGL7 PARTE	Q6bgl7 paramecium
34	101.5	13.6	556	2	Q9V7I9 DROME	Q9v7i9 drosophila
35	101.5	13.6	674	2	Q7RLE7 PLAYO	Q7rle7 plasmodium
36	101.5	13.6	785	2	Q9GQ82 DROME	Q9gq82 drosophila
37	101.5	13.6	954	2	Q6HNR0 BACHK	Q6hnr0 bacillus th
38	101	13.5	329	2	Q9NFV9 PLAPA	Q9nf9 plasmodium
39	100.5	13.4	1015	2	Q5W5T1 TETPY	Q5w5t1 tetrahymena
40	100	13.4	312	2	Q9PPL5 CANJE	Q9ppl5 campylobact
41	100	13.4	400	2	Q5AL15 CANAL	Q5al15 candida alb
42	100	13.4	1011	2	Q4Y2I3 PLACH	Q4y2i3 plasmodium
43	100	13.4	1130	2	Q81JZ4 PLAPF7	Q81jz4 plasmodium
44	100	13.4	1550	2	Q54GS1 DICDI	Q54ge1 dictyosteli
45	99.5	13.3	467	2	Q59PR2 CANAL	Q59pe2 candida alb

#### ALIGNMENTS

RESULT 1  
Q9AHT5 STRPN PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Serine protease (Fragment).  
GN Name=prtA;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
CC Streptococcus.  
OX NCBI\_Taxid=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
RX DOI=10.1128/JAI.69.3.1593-1598.2001;  
RA Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,  
RA Langermann S., Johnson S., Koenig S.;  
RT "Use of a whole genome approach to identify vaccine molecules  
affording protection against Streptococcus pneumoniae infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
DR EMBL; AF291699; AAK19159.1; -; Genomic\_DNA.  
DR HSSP; P00782; 288T.  
DR MEROPS; S08.064; -;  
DR GO; GO:0009986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0004289; F:subtilase activity; IEA.  
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; Subtilisin.  
DR TIGRfams; TIGR01167; LPTXG anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Protease.

```

FT NON TER 1 1
SQ SEQUENCE 2119 AA; 238227 MW; 517F9B7F6B960A6A CRC64;

Query Match 100.0%; Score 748; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 4.7e-46;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRVTVTIQKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVYINLSK 60
|||||
Db 1942 HRVTVTIQKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVYINLSK 2001
|||||

Qy 61 DTFIKPVFKIEEKKKEENKPTFDVSKKONPVNHSQNLNESHKREDLQREHSQKSDST 120
|||||
Db 2002 DTFIKPVFKIEEKKKEENKPTFDVSKKONPVNHSQNLNESHKREDLQREHSQKSDST 2061
|||||

Qy 121 KDVTTATVLDKNNISSKSTTNPNK 144
|||||
Db 2062 KDVTTATVLDKNNISSKSTTNPNK 2085
|||||

RESULT 2
Q9TRY6_STRPN
ID Q9TRY6_STRPN PRELIMINARY; PRT; 2140 AA.
AC Q9TRY6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Serine protease, subtilase family.
GN OrderedLocuNames=SP0641;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]_TaxID=1313;
RP NUCLEOTIDE SEQUENCE.
RC STRAINS-ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.P., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer C.M.;
RA "Complete genome sequence of a virulent isolate of Streptococcus
RA pneumoniae.";
RT Science 293:498-506(2001).
RL EMBL; AE007373; AAK74791.1; -; Genomic_DNA.
DR PIR; F95074; F95074.
DR HSSP; P00782; 2S8T.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:00043086; F:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin N; 1.
DR PRINTS; PR00723; SUBTILISIN.

```

```

DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

Query Match 100.0%; Score 748; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 4.8e-46;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRVTVTIQKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVYINLSK 60
|||||
Db 1963 HRVTVTIQKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGVYINLSK 2022
|||||

Qy 61 DTFIKPVFKIEEKKKEENKPTFDVSKKONPVNHSQNLNESHKREDLQREHSQKSDST 120
|||||
Db 2023 DTFIKPVFKIEEKKKEENKPTFDVSKKONPVNHSQNLNESHKREDLQREHSQKSDST 2082
|||||

Qy 121 KDVTTATVLDKNNISSKSTTNPNK 144
|||||
Db 2083 KDVTTATVLDKNNISSKSTTNPNK 2106
|||||

RESULT 3
Q8DQ7_STRR6
ID Q8DQ7_STRR6 PRELIMINARY; PRT; 2144 AA.
AC Q8DQ7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN Name=prtA; OrderedLocuNames=spr0561;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]_TaxID=171101;
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Albom W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rosteck P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RA "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RT J. Bacteriol. 183:5709-5717(2001).
RL EMBL; AE008434; AAK99365.1; -; Genomic_DNA.
DR PIR; A97942; A97942.
DR HSSP; P00782; 2S8T.
DR MEROPS; S08.064; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.

```

DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Complete proteome.  
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.6%; Score 745; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 7.9e-46;  
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWIEISGFEGKKDAGVYINLSK 60  
DB 1967 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWIEISGFEGKKDAGVYINLSK 2026

QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSHQKSDST 120  
DB 2027 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSHQKSDST 2086

QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
DB 2087 KDVTATVLDKNNISSKSTNNPNK 2110

RESULT 4  
Q9S4M8 STRPN PRELIMINARY; PRT; 2144 AA.  
AC Q9S4M8;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAY-2004 (TrEMBLrel. 26, Last annotation update)  
DE Cell wall-associated serine proteinase precursor PrtA.  
GN Name=prtA;  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=3.B;  
RX MEDLINE=21585565; PubMed=11728722;  
RA Bethe G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,  
RA Zysk G.;  
RT "The cell wall-associated serine proteinase PrtA: a highly conserved  
RT virulence factor of Streptococcus pneumoniae.";  
RL FEMS Microbiol. Lett. 205:99-104(2001).  
DR EMBL; AF127143; AAD48399.1; -; Genomic\_DNA.  
DR HSP; P00782; 2SPT.  
DR MEROPS; S08.064; -.  
DR GO; GO:000986; C:cell surface; IEA.  
DR GO; GO:0005618; C:cell wall; IEA.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0008233; F:peptidase activity; IEA.  
DR GO; GO:0042802; F:protein self binding; IEA.  
DR GO; GO:0042893; F:subtilase activity; IEA.  
DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.  
DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR010435; DUF1034.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Pept\_S8\_S53.  
DR InterPro; IPR010259; Prot\_inh\_S8A.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF06280; DUF1034; 1.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 1.  
DR Pfam; PF05922; Subtilisin\_N; 1.

DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; UNKNOWN 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; UNKNOWN 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; UNKNOWN 1.  
KW Cell wall; Signal.  
FT SIGNAL 1 19 Potential.  
FT CHAIN 20 2144 cell wall-associated serine proteinase  
FT PrtA.  
SQ SEQUENCE 2144 AA; 240725 MW; 2052511470741331 CRC64;

Query Match 99.5%; Score 744; DB 2; Length 2144;  
Best Local Similarity 98.6%; Pred. No. 9.4e-46;  
Matches 142; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWIEISGFEGKKDAGVYINLSK 60  
DB 1967 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWIEISGFEGKKDAGVYINLSK 2026

QY 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSHQKSDST 120  
DB 2027 DTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSHQKSDST 2086

QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
DB 2087 KDVTATVLDKNNISSKSTNNPNK 2110

RESULT 5  
Q4XUI6 PLACH PRELIMINARY; PRT; 300 AA.  
AC Q4XUI6;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Hypothetical protein (Fragment).  
GN ORFNames=PC000286.03.0;  
OS Plasmodium chabaudi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
OX NCBI\_TaxID=5825;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolij T.W.A.,  
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,  
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,  
RA Quail M.A., Ormond D., Doggett J., Trueman H.B., Mendoza J.,  
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,  
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;  
RT "A comprehensive survey of the Plasmodium life cycle by genomic,  
RT transcriptomic, and proteomic analyses.";  
RL Science 307:82-86(2005).  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; CAJ701003049; CAH79425.1; -; Genomic\_DNA.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 4.  
DR PRINTS; PR00320; GPROTEINRPT.  
DR PRODOM; PD0000016; WD40; 3.  
DR SMART; SM00330; WD40; 5.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
DR PROSITE; PS00082; WD\_REPEATS\_2; 4.  
DR PROSITE; PS02094; WD\_REPEATS\_REGION; 1.  
KW Hypothetical protein; Repeat; WD repeat.  
FT NON\_TER 1  
SQ SEQUENCE 300 AA; 34469 MW; 8ED4E512AFB1945 CRC64;

Query Match 17.2%; Score 128.5; DB 2; Length 300;  
Best Local Similarity 26.6%; Pred. No. 0.15;  
Matches 37; Conservative 31; Mismatches 38; Indels 33; Gaps 7;

QY 9 NGKMSSTIVSE-----EDFILPVYKGELEKGYQFDGWIEISGFEGKKDAGVYINLSK 60

Db 140 NGKFIASSSFDKSIWISGIDGTLYAVRGHVGPAYKI-ANSI-----DNNYIVSCSQ 191  
Qy 61 DTFPIK-----PVPKLEEKKE---BENKPTFDVSKKDNQPNVNHSQLNESHKEDLQ 109  
Db 192 DSTLKLWINHLVPLLLKKKEBNAEQTKDEQK-----NEQKENPQ-NNQPNDNEANSEBKK 245  
Qy 110 REHSQKSDSTKDVATVTL 128  
Db 246 KKEKEKNDKTKNKIKTLL 264

## RESULT 6

ID Q95P15\_PLAFA PRELIMINARY; PRT; 361 AA.  
AC Q95P15;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=FVO;  
RX MEDLINE=21853556; PubMed=11865423; DOI=10.1086/339187;  
RA Hiseida H., Saul A., Reese J.J., Kennedy M.C., Long C.A., Miller L.H.,  
RA Stowers A.W.;  
RT "Merozoite surface protein 3 and protection against malaria in Aotus  
nancymai monkeys";  
RL J. Infect. Dis. 185:657-664 (2002).  
DR EMBL: AY044180; AAK94780.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
KW Merozoite.  
FT CHAIN <1 >361 merozoite surface protein 3.  
FT NON\_TER 1  
FT NON\_TER 361  
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 15.0%; Score 112; DB 2; Length 361;

Best Local Similarity 23.9%; Pred. No. 3;

Matches 37; Conservative 28; Mismatches 48; Indels 42; Gaps 6;

Qy 28 YKGELEKGYQ-----PD-----GWEISGF--EGKKDAG-----YVI 56  
Db 165 YAGKVEKDYERAKNAYQKANQAVLKAKEASSYDYLGWFGGVPHEKKENMLSHLYVS 224  
Qy 57 NLSKDTFTKPVFKIEEKKEENKPTFDVSKKDNQPNVNHSQLNESHKEDLQREHSQK 116  
Db 225 SKDKENISKENDVDLDE-KEEAEETEEELKEKNEETESISEDEDEEEEEEKEE 283  
Qy 117 SDSTKDVATVLDKN-----NISKSTNN 141  
Db 284 NDKKKEQEKQSNENNNDQKMEAGNLISKQNNN 318

## RESULT 7

ID Q9U0G0\_PLARE PRELIMINARY; PRT; 346 AA.  
AC Q9U0G0;  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
GN Name=msp3;  
OS Plasmodium reichenowi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5854;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;

RA Okeno D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
Plasmodium reichenowi and Plasmodium falciparum";  
RL Mol. Biochem. Parasitol. 109:185-188 (2000).  
DR EMBL: AJ252286; CAB65754.1; -; Genomic\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
KW Merozoite.  
FT NON\_TER 1  
FT NON\_TER 346  
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAPA010 CRC64;

Query Match 14.7%; Score 110; DB 2; Length 346;

Best Local Similarity 26.2%; Pred. No. 4;

Matches 34; Conservative 26; Mismatches 54; Indels 16; Gaps 5;

Qy 27 VYKGELEKGYQFD-GWEISGF--EGKKDAG-----YVINLSKDTFTKPVFKIEEKKEE 78  
Db 184 VLKAKEASSYNYILGWFGGVPHEKKENMLSHLYVSKDKENISKENDVDLDE-KEE 242  
Qy 79 NKPTFDVSKKDNQPNVNHSQLNESHKEDLQREHSQKSDSTKDVATV-----LDKN 131  
Db 243 ABETGQELEKNEETESISEDEDEEEEEEKEENNDQKMEAGNLISKQNNN 302  
Qy 132 NISKSTNN 141  
Db 303 NLISKQNNN 312

## RESULT 8

ID Q25705\_PLAFA PRELIMINARY; PRT; 379 AA.  
AC Q25705;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;  
RA McColl D.J., Anders R.P.;  
RT "Conservation of structural motifs and antigenic diversity in the  
Plasmodium falciparum merozoite surface protein-3 (MSP-3)";  
RL Mol. Biochem. Parasitol. 90:21-31 (1997).  
DR EMBL: U08851; AAC47831.1; -; Unassigned\_DNA.  
DR InterPro; IPR010784; Merozoite\_SPAM.  
DR Pfam; PF07133; Merozoite\_SPAM; 1.  
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 14.7%; Score 110; DB 2; Length 379;

Best Local Similarity 23.4%; Pred. No. 4.4;

Matches 37; Conservative 30; Mismatches 41; Indels 50; Gaps 7;

Qy 28 YKGELEKGYQ-----PD-----GWEISGF--EGKKDAG-----YVI 56  
Db 185 YAGKVEKDYERAKNAYQKANQAVLKAKEASSYDYLGWFGGVPHEKKENMLSHLYVS 244  
Qy 57 NLSKDTFTKPVFKIEEKKEENKPTFDVSKKDNQPNVNHSQLN-----ESH 103  
Db 245 SKDKENISKENDVDLDE-KEEAEETEEELKEKNEETESISEDEDEEEEEEKEE 303  
Qy 104 RKEDLQREHSQKSDSTKDVATVLDKNISKSTNN 141  
Db 304 KKEQKEQESNENNNDQKMEAGNLISKQNNN 336

## RESULT 9

Q8CPK8\_STAEP PRELIMINARY; PRT; 775 AA.  
ID Q8CPK8\_STAEP PRELIMINARY; PRT; 775 AA.  
AC Q8CPK8;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Penicillin-binding protein 1.  
 GN OrderedLocusNames=SE0856;  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=1282;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=ATCC 12228;  
 RX PubMed=1295022; DOI=10.1046/j.1365-2958.2003.03671.x;  
 RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,  
 RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,  
 RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;  
 RT "Genome-based analysis of virulence genes in a non-biofilm-forming  
 ST Staphylococcus epidermidis strain (ATCC 12228).";  
 RL Mol. Microbiol. 49:1577-1593 (2003).  
 DR EMBL: AE016746; AA004453.1; -; Genomic\_DNA.  
 DR HSP; P14677; 1QME.  
 DR GO: GO:0008659; P: penicillin binding; IEA.  
 DR GO: GO:0009273; P: cell wall biosynthesis (sensu Bacteria); IEA.  
 DR InterPro: IPR005543; PASTA.  
 DR InterPro: IPR005311; PBP dimer.  
 DR InterPro: IPR012338; PBP\_trypt\_fold.  
 DR InterPro: IPR01460; Pencil\_bind\_trypt.  
 DR Pfam: PF03793; PASTA; 2.  
 DR Pfam: PF03717; PBP dimer; 1.  
 DR Pfam: PF00905; Transpeptidase; 1.  
 DR SMART: SM00740; PASTA; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;

Query Match 14.7%; Score 110; DB 2; Length 775;  
 Best Local Similarity 27.0%; Pred. No. 9.5;  
 Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;

Qy 5 VTIONGKMSSTIVSEEDFILPVYK-----GELEKGYQPDGW---EISGFE-----G 48  
 Db 625 ITIGNGKIQKQSVKSGTKVLPHSKVLMVTDGELTMP-DMTGWTKEVDLAPEDLTIKVVS 683

Qy 49 KKDAGYVNL--LSKDTFIPKPKLEKKEENKPTDVS-----KKKNPQVNSQLNES 102  
 Db 684 TKGNGFVNTQSGISGQIIK-----NKKIEVLSAEDTDDQSKTDESDSDN 730

Qy 103 HRKEDLQREHSQKSDSTKVATVLDKNKISSKSTNN 141  
 Db 731 KSKDKADEHSNTSSSTKN-----DKSNADSKNDSDD 763

RESULT 10  
 Q81436 PLAF7  
 ID Q81436 PLAF7 PRELIMINARY; PRT; 3008 AA.  
 AC Q81436  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Hypothetical protein PPE0325w.  
 GN Name=PPE0325w;  
 OS Plasmodium falci-parum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornou N.,  
 RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford T., Holroyd S., Horrocks P.,

RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford T., Holroyd S., Horrocks P.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;  
 RT "Sequence of Plasmodium falci-parum chromosomes 1, 3-9 and 13.";  
 RL Nature 419:527-531 (2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=3D7;  
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;  
 RL Submitted (SSP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL929351; CAD51431.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 3008 AA; 356025 MW; 60BCBBEE15C599B4 CRC64;

Query Match 14.6%; Score 109; DB 2; Length 3008;  
 Best Local Similarity 32.4%; Pred. No. 48;  
 Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

Qy 54 YVNLK-----DTFIKVPFKIKBEKKEENKPTDVSKKONPQVNSIQ-----LNEHSRKE 106  
 Db 2310 YDIELSKIERFGASIGVFTD-BENKBEKN--EVNKKBEKBEKBEKBEKBEK 2366

Qy 107 DLQREH-----SOKSDSTKVATVLDKNKISSK-----STNNPNK 144  
 Db 2367 ENKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEKBEK 2414

RESULT 11  
 O77355 PLAF7  
 ID O77355 PLAF7 PRELIMINARY; PRT; 600 AA.  
 AC O77355;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein WAL3P4.20.  
 GN Name=WAL3P4.20; Synonyms=PFC0465c;  
 OS Plasmodium falci-parum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;  
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,  
 RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
 RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,  
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,  
 RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;  
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium  
 falci-parum.";  
 RL Nature 400:532-538 (1999).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornou N.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford T., Holroyd S., Horrocks P.,

[illegible]

```

Qy 102 SHRKEDLQREHSHQKSDSTKD--VTATVLDKNNISSK 136
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 324 SEKBESKKEKHSHKEDKKKDEBKSKKVEDKSKQK 360

RESULT 16
Q815F3_PLAF7 Q815F3_PLAF7 PRELIMINARY; PRT; 829 AA.
AC Q815F3;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PFL1275G;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255705; PubMed=12368664; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyse S.,
RA Chan M.-S., Nene V., Shallow J.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramaniam G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Praser C.M., Barrell B.G.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
RN [2];
RP NUCLEOTIDE SEQUENCE.
RA Hyman R.W., Fung E., Conway A., Kurdi O., Mao J., Miranda M.,
RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF014846; AAN36341.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 829 AA; 98815 MW; EF2675E301B2CE93 CRC64;

Query Match 14.3%; Score 107; DB 2; Length 829;
Best Local Similarity 24.8%; Pred. No. 17;
Matches 36; Conservative 30; Mismatches 43; Indels 36; Gaps 6

Qy 16 TIVSEDFILPVYKGELEKGYQFGWISGPEGKIDAGYINLSKDTFIKVPFKIEEK 75
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 519 TLMTERTDVLDPVTHPMYKKYHLN-----KGTLPNFSLN-----PRIEKKN 561

Qy 76 EEENKPT-----FDVSKKKDN--POVHNSQLNESH-RKEDLQREHSHQKSDS 119
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 562 RDKKKGTHNNKNDAEYMLKYKKKKKKQFEENNTELNDNSIKKNNKLVEH---DNS 618

Qy 120 TKDVTATVLDKNNISSKSTTNPNK 144
      |:::|:::|:::|:::|:::|:::|:::|:::|
Db 619 LKQEQLIINDKNVIEHTKIYDNQKK 643

RESULT 17
Q5V9M0_PLAKN Q5V9M0_PLAKN PRELIMINARY; PRT; 374 AA.
AC Q5V9M0;
DT 01-FEB-2005 (TREMBLrel. 29, Created)
DT 01-FEB-2005 (TREMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TREMBLrel. 29, Last annotation update)
DE Merozoite surface protein 5.
GN Name=MSP5;
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=H;
```

```
RA Black C.G., Wang L., Topolska A.E., Finkelstein D.I., Horne M.K.,
RA Thomas A.W., Mohandas N., Coppel R.L.;
RT "Merozoite surface proteins 4 and 5 of Plasmodium knowlesi have
RT differing cellular localisation and association with lipid rafts.";
DR MBL; Biochem. Parasitol. 138:153-158(2004).
DR EMBL; AY573058; AFAT5729.1; -; Genomic_DNA.
DR InterPro; IPR006209; EGF_like.
KW Merozoite; EGF; 1.
SQ SEQUENCE 374 AA; 41642 MW; C0D687C6F23FE989 CRC64;

Query Match 14.2%; Score 106.5; DB 2; Length 374;
Best Local Similarity 25.8%; Pred. No. 7.8;
Matches 39; Conservative 28; Mismatches 63; Indels 21; Gaps 6;

QY 8 QNGKMSSTIVSEEDFILPVYKGL--RKG-----YQFDGWEISFGKQAGYVINL-- 58
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 19 QSGHPLKFSFWREERKIHLYTNRLREBEGKNDVQEMESPISGTEGKGKIQMTISHLQL 78
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 59 -----SKDTFKPVFKTBEKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQREEH 113
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 79 QSGKIDQVSLFANDSNLKGANEASGSGKSGDEENVKUSD-BENAKSD--BEN 134
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 114 SQKSDSTKDVATVLDKNNISSK---TTNN 141
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 135 KDANGNTKDAESAEGEENPVSQENQMKTLNN 165
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 18
Q9FJK9 ARATH PRELIMINARY; PRT; 470 AA.
AC Q9FJK9;
DT 01-WAR-2001 (TrEMBLrel. 16, Created)
DT 01-WAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Gb(AAF20218.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP MEDLINE=99087489; PubMed=9872454;
RX Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
DR EMBL; AB015468; BAB10694.1; -; Genomic DNA.
SQ SEQUENCE 470 AA; 53758 MW; 6D866CE72E35AC54 CRC64;

Query Match 14.2%; Score 106.5; DB 2; Length 470;
Best Local Similarity 20.3%; Pred. No. 10;
Matches 36; Conservative 31; Mismatches 65; Indels 45; Gaps 5;

QY 1 HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKYQFDGWEISGFE-----GKDA 52
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 84 NRVTDTVQNNSGSK-----YVQDLARRIRYDE-BATGSAQRIDHPNQNV 131
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 53 GYVINLSKDTFTKPVFKTBEKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQREEH 193
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 132 GITEKAPENSPIETFSHRVDDNKRINNQNFTAAKSSSENAVSRVSGADHKKAQVNGKPM 191
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 VNHSQINE-----SHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPK 144
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 192 ENRDQVRQTESAEKSHRKENVTKSEKPRDQGVKKTEAKDKDRNKEKKEKTESINK 248
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 19
Q6BRW2 DEBHA PRELIMINARY; PRT; 616 AA.
ID Q6BRW2 DEBHA PRELIMINARY; PRT; 616 AA.
```

```
AC Q6BRW2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to CH4458|IPF8464 Candida albicans IPF8464 unknown
DE function.
GN OrderedLocusNames=DEBHA0D14674g;
OS Debaryomyces hansenii (Yeast) (Torulasporea hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36219 / CBS 767;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boisarame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Pukhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,
RT Wincker P., Souciet J.-L.;
RL "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR382136; CAG87226.1; -; Genomic DNA.
DR GO; GO:0016301; P:kinase activity; IEA.
DR InterPro; IPR000749; ATP-gua_Ptrans.
KW Complete proteome.
SQ SEQUENCE 616 AA; 72143 MW; 884009B28B86C3CF CRC64;

Query Match 14.2%; Score 106; DB 2; Length 616;
Best Local Similarity 30.7%; Pred. No. 15;
Matches 43; Conservative 20; Mismatches 49; Indels 28; Gaps 9;

QY 26 PVYKGELEKGYQFDGWEISFGKQAGYVINLSKDT-FIKPV----FKKIEEKKEENK 80
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 87 PVLKGRRAKFKITLDTIT-----KDIN-DINFSDSSEHEKPIETSKTKETKTGTK 140
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 PTFDVSKKK-----DNFQ--VNHSQINE-----SHRKEDLQREHSQK-----SDSTKDV 124
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 141 PDLDIGKLERIRVADNPDEIENHSSKSEIKQRREKQRQRKRLKRLKANQESNNDS 200
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 125 ATVLDK-NNISSKSTTNNPN 143
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 201 TEQPEPLKINKEKITSNEPS 220
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 20
Q9U6C4 PLAPA PRELIMINARY; PRT; 379 AA.
ID Q9U6C4 PLAPA PRELIMINARY; PRT; 379 AA.
AC Q9U6C4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Polymorphic antigen.
GN Name=MSP-3;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PCCL/EN;
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Pang J.M., Xu J.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF188190; AAF04099.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
```



```
DR Pfam: PF07133; Merozoite_SPAM; 1.
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1P9DSF25 CRC64;

Query Match      14.0%; Score 104.5; DB 2; Length 379;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

QY 7 IQNGKEMSTIVSEEDFILPVYKGLKGYQPDGWEISGF--EGKKDAG-----YVINLS 59
Db 207 VLKAKEASS-----YDYL-----GWFGGVPPEHKKEENMLSHLYVSSKD 247
QY 60 KDTFTKPVFKLKEEKEENKPTFVSKKONPQVNHSLN-----ESHRKE 106
Db 248 KENISKENDVDLDE--KEEAEETEKEEKEENKPTFVSKKONPQVNHSLN-----ESHRKE 106
QY 107 DLQREHQSOKSDTKDVTATVLDDKNNISKSTNN 141
Db 307 EQEKEQSNNNDQKKDMEA-----QNLISKQNNN 336

RESULT 21
Q26019 PLAFPA PRELIMINARY; PRT; 380 AA.
AC Q26019_
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Polymorphic antigen precursor.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95198774; PubMed=7891748; DOI=10.1016/0166-6851(94)00149-9;
RC STRAIN=FC27;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favalaro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.P.;
RT "Molecular variation in a novel polymorphic antigen associated with
RT Plasmodium falciparum merozoites.";
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FC27;
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L07944; AAC09378.1; -; Genomic_DNA.
DR PDB; 1FSM; NMR; @=90-127.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite_SPAM; 1.
KW Signal.
FT SIGNAL. 1 25 Potential.
FT CHAIN 26 380 polymorphic antigen.
SQ SEQUENCE 380 AA; 43290 MW; 0986CA139309CA2 CRC64;

Query Match      14.0%; Score 104.5; DB 2; Length 380;
Best Local Similarity 24.5%; Pred. No. 11;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

QY 7 IQNGKEMSTIVSEEDFILPVYKGLKGYQPDGWEISGF--EGKKDAG-----YVINLS 59
Db 208 VLKAKEASS-----YDYL-----GWFGGVPPEHKKEENMLSHLYVSSKD 248
QY 60 KDTFTKPVFKLKEEKEENKPTFVSKKONPQVNHSLN-----ESHRKE 106
Db 249 KENISKENDVDLDE--KEEAEETEKEEKEENKPTFVSKKONPQVNHSLN-----ESHRKE 106
QY 107 DLQREHQSOKSDTKDVTATVLDDKNNISKSTNN 141
Db 308 EQEKEQSNNNDQKKDMEA-----QNLISKQNNN 337
```

```
RESULT 22
Q4YVY2 PLABE PRELIMINARY; PRT; 736 AA.
ID Q4YVY2_
AC Q4YVY2_
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=PB000556.02.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolj T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
KW Hypothetical protein.
DR EMBL; CAI01002190; CAH97824.1; -; Genomic_DNA.
SQ SEQUENCE 736 AA; 85816 MW; A1315CF2D97A6905 CRC64;

Query Match      14.0%; Score 104.5; DB 2; Length 736;
Best Local Similarity 23.0%; Pred. No. 23;
Matches 41; Conservative 34; Mismatches 60; Indels 43; Gaps 8;

QY 7 IQNGKEMSTIVSEEDFILPVYKGLKGYQPDGWEISGFEGKKDAGVINLSKDT----- 62
Db 401 INNNDSSS---GKNNAIPIFLKCKKNNKCDTIETNNPWINKAGDVLKCKKESLTLL 457
QY 63 PIKPVFKKIEKKEE-----ENKPTFVSKKONPQVNH-----SQLNESH----- 104
Db 458 FSSPLHKKIEKKEVINDEYNAKVENIKKNNKGRIKHPLKIEFYFNEKINTPKPPDD 517
QY 105 ----KEDLQREHQSOKSD-----STKDVAT--VLDK-----NNTSSKSTNNPK 144
Db 518 IISANENFKTNSPFGKSDILDTSVNNINANYEIEFGKDKIKNSNSKVDSENIIDNDK 575

RESULT 23
Q4YMU4 PLABE PRELIMINARY; PRT; 662 AA.
ID Q4YMU4_
AC Q4YMU4_
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE RNA binding protein, putative.
GN ORFNames=PB001104.03.0;
OS Plasmodium berghei.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5821;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Hall N., Karras M., Raine J.D., Carlton J.M., Koolj T.W.A.,
RA Berriman M., Florens L., Janssen C.S., Pain A., Christophides G.K.,
RA James K., Rutherford K., Harris B., Harris D., Churcher C.,
RA Quail M.A., Ormond D., Doggett J., Trueman H.E., Mendoza J.,
RA Bidwell S.L., Rajandream M.A., Carucci D.J., Yates J.R., Kafatos F.C.,
RA Janse C.J., Barrell B., Turner C.M.R., Waters A.P., Sinden R.S.;
RT "A comprehensive survey of the Plasmodium life cycle by genomic,
RT transcriptomic, and proteomic analyses.";
RL Science 307:82-86(2005).
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
```



```
AC Q25706;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Polymorphic antigen.
OS Plasmodium falci-parum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98156743; PubMed=9497029; DOI=10.1016/S0166-6851(97)00130-8;
RA McColl D.J., Anders R.P.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falci-parum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08852; AAC47832.1; -; Unassigned_DNA.
DR InterPro; IPR010784; Merozoite SPAM.
DR Pfam; PF07133; Merozoite SPAM; 1.
SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;

Query Match 13.8%; Score 103.5; DB 2; Length 379;
Best Local Similarity 24.5%; Pred. No. 13;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

QY 7 IQNGKEMSTIVSEEDFILPVYKGLKGYQFDGWEISGP--EGKKDAG-----YVINLS 59
DB 207 VLKAEASG-----YDYL-----GWFGGVPFHKKEKRNMLSHLYVSKD 247
QY 60 KDTFTKPVFKIEEKKENKPTFDVSKKKNPQVNHSLN-----ESHRKE 106
DB 248 KENISKENDVDLDE-KEEAETETEELEKEKVEETSEISEDEBEREEREKENDKK 306
QY 107 DLQREHESQKSDTKDVTATVLDKNISKSTNN 141
DB 307 EQEKEQSNENNQKKOMEA-----QNLISKQNNN 336

RESULT 27
Q50LX8 ENTHI PRELIMINARY; PRT; 296 AA.
AC Q50LX8;
DT 13-SEP-2005 (TReMBLrel. 31, Created)
DT 13-SEP-2005 (TReMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TReMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=657.t00001;
OS Entamoeba histolytica HM-1-IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Almark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich S., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoft U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagals K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitz S., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gluchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sicheritz-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
KW Hypothetical protein.
SQ SEQUENCE 296 AA; 33757 MW; 3A5986BB34A7FC3B CRC64;
```

```
Query Match 13.8%; Score 103; DB 2; Length 296;
Best Local Similarity 28.4%; Pred. No. 11;
Matches 40; Conservative 23; Mismatches 50; Indels 28; Gaps 6;

QY 11 KEMSSTIVSEEDFILPVYKGLKGYQFDGWEISGFEKKDAGYVINLSKDTFTKPVFK 70
DB 135 KEQKTKESGE-----GDSEKKH-----DPTNEGKKN-----DTTKDNKDKEKD 177
QY 71 IEKKK---EENKPTFDVSKKK-DNPQVNH-----QLNESHKRLDLQREHESQKSDS 119
DB 178 TNEBEGSGKGQKTNBESKKTENPQNNESNKEEKKKEEKKKEEKKKEEKKKEE 237
QY 120 TKDVTATVLDKNISKSTNN 140
DB 238 EQNEKFPINEKNEKENKNTN 258
DB 238 EQNEKFPINEKNEKENKNTN 258

RESULT 28
Q5HQ11 STABQ PRELIMINARY; PRT; 775 AA.
AC Q5HQ11;
DT 10-MAY-2005 (TReMBLrel. 30, Created)
DT 10-MAY-2005 (TReMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TReMBLrel. 30, Last annotation update)
DE Penicillin-binding protein 1.
GN Name=pbpi; OrderedLocNames=SERP0746;
OS Staphylococcus epidermidis (strain ATCC 35984 / RP62A).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=176279;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15774866; DOI=10.1128/JB.187.7.2426-2438.2005;
RA Gill S.R., Fouts D.E., Archer G.L., Mongodin E.F., DeBoy R.T.,
RA Ravel J., Paulsen I.T., Kolonay J.F., Brinkac L.M., Beanan M.J.,
RA Dodson R.J., Daugherty S.C., Madupu R., Angiuoli S.V., Durkin A.S.,
RA Haft D.H., Vamathevan J.J., Khouri H., Utterback T.R., Lee C.,
RA Dimitrov G., Jiang L., Qin H., Weidman J., Tran K., Kang K.H.,
RA Hance I.R., Nelson K.E., Fraser C.M.;
RT "Insights on evolution of virulence and resistance from the complete
RT genome analysis of an early methicillin-resistant Staphylococcus
RT aureus strain and a biofilm-producing methicillin-resistant
RT Staphylococcus epidermidis strain.";
RL J. Bacteriol. 187:2426-2438(2005).
DR EMBL; CP000029; AAM54126.1; -; Genomic_DNA.
DR TIGR; SERP0746; -;
DR GO; GO:0008658; P:penicillin binding; IEA.
DR GO; GO:0009273; P:cell wall biosynthesis (sensu Bacteria); IEA.
DR InterPro; IPR005543; PASTA.
DR InterPro; IPR005311; PBP_dimer.
DR Pfam; PF03793; PASTA; 2.
DR Pfam; PF03717; PBP_dimer; 1.
DR Pfam; PF00905; Transpeptidase; 1.
DR SMART; SM00740; PASTA; 2.
DR Complete proteome.
SQ SEQUENCE 775 AA; 86352 MW; B9395893E0043694 CRC64;

Query Match 13.7%; Score 102.5; DB 2; Length 775;
Best Local Similarity 24.7%; Pred. No. 34;
Matches 38; Conservative 26; Mismatches 61; Indels 29; Gaps 7;

QY 5 VTIQNGKEMSTIVSEEDFILPVYK-----GELEKGYQFDGW---EISGFE-----G 48
DB 625 ITIGNGKQIKQQSVKSGTKVPLPHSKVLMWTDGLTMP-DMTGWTKEVDLAFEDLTIKVVS 683
QY 49 KKDAGYVIN--LSKDTFTKPVFK-----KIEKKKEENKPTFDVSKKKNPQVNHQS 98
DB 684 TKGNFVTNQSIKSGQIKKDKTEVSLSAEDTDDQKEDTDESSDKSKKDKVDENSN 743
QY 99 LNEHSHRKEDLQREHESQKSDTKDVTATVLDKNN 132
DB 744 ASSSKKNEKSNADSKNDSDDSTNETSGS--ERNN 775
```

```

RESULT 29
Q512T7 ENTHI PRELIMINARY; PRT; 1069 AA.
AC Q512T7;
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Fimbriae-associated protein, putative.
GN ORFNames=89.t00007;
OS Entamoeba histolytica HM-1:IMSS.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=294381;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HM-1:IMSS;
RX PubMed=15729342; DOI=10.1038/nature03291;
RA Loftus B., Anderson I., Davies R., Alemark U.C., Samuelson J.,
RA Amedeo P., Roncaglia P., Berriman M., Hirt R.P., Mann B.J., Nozaki T.,
RA Suh B., Pop M., Duchene M., Ackers J., Tannich B., Leippe M.,
RA Hofer M., Bruchhaus I., Willhoest U., Bhattacharya A.,
RA Chillingworth T., Churcher C., Hance Z., Harris B., Harris D.,
RA Jagels K., Moule S., Mungall K., Ormond D., Squares R., Whitehead S.,
RA Quail M.A., Rabinowitsch E., Norbertczak H., Price C., Wang Z.,
RA Guillen N., Gilchrist C., Stroup S.E., Bhattacharya S., Lohia A.,
RA Foster P.G., Sichert-Ponten T., Weber C., Singh U., Mukherjee C.,
RA El-Sayed N.M., Petri W.A., Clark C.G., Embley T.M., Barrell B.,
RA Fraser C.M., Hall N.;
RT "The genome of the protist parasite Entamoeba histolytica.";
RL Nature 433:865-868 (2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAF01000328; EAL47849.1; -; Genomic DNA.
SQ SEQUENCE 1069 AA; 120249 MW; 1D38E4F0A8759CB4 CRC64;

Query Match 13.7%; Score 102.5; DB 2; Length 1069;
Best Local Similarity 25.5%; Pred. No. 47;
Matches 39; Conservative 31; Mismatches 54; Indels 29; Gaps 7;

QY 6 TTIONG-KEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKDGAGYVNLNSKDTFI 64
Db 150 TLTNGEKKLSTSLCNEQD-----ELQKSSSSSTD-NKNDKRDRIHFVDVLPKNEE 199

QY 65 KVPFKKIEEKEENK-----PTFDVSKKKNQ-----VNHSLNESHRKEDLQREE 112
Db 200 KEISWEIESKTEERKSNLQIPSLNSLSEKDKNESVEIAKVLKKSNNSSNGEEDKQDEE 259

QY 113 HS-QKSDSTKD-----VTATVLDKNNISSKSTT 139
Db 260 VSEKFDSEQEEKKEEMIKAEVSQNKVKDKSTT 292

RESULT 30
Q9LH98 ARATH PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone: t19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
OC rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20363099; PubMed=10907853;
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
RT and BAC clones.";
```

```

RL DNA Res. 7:217-221(2000).
RN (2)
RP NUCLEOTIDE SEQUENCE.
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL EMBL; AP002057; BAB03174.1; -; Genomic DNA.
DR HSSP; P01096; 1HP9.
DR InterPro; IPR009605; DUF1216.
DR Pfam; PF06746; DUF1216; 2.
SQ SEQUENCE 2081 AA; 232852 MW; L3603ELP8SEFFP29 CRC64;

Query Match 13.7%; Score 102.5; DB 2; Length 2081;
Best Local Similarity 27.4%; Pred. No. 97;
Matches 43; Conservative 24; Mismatches 63; Indels 27; Gaps 6;

QY 4 TTTIONGKEMSSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFGKDK-----AG 53
Db 1659 TVEINGGEELSTERGSKD-----GKIBEGK--EGKENSTKEGSKDKIEEGMEGKEN 1708

QY 54 YVINLSKDTKTPKVPFKIEEKKBEENK--TFDVS-K-KDNQPVNHSQLNESHKEDLQ- 109
Db 1709 STKESSKDGKINETHGDKKATMERGSKDGGTNGTSKSKSKSVKVEINGVDKDSLKDSKN 1769

QY 110 ---REHSQKSDSTKQVATVLDKNNISSKSTTNNPN 143
Db 1769 GDINEINGKEDSVKDNVTIQGNDNSLTNSTSEPN 1805

RESULT 31
Q54WT2 DICDI PRELIMINARY; PRT; 540 AA.
AC Q54WT2;
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=DD0186654;
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AX4;
RA Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,
RA Sugchang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA Pilcher K., Chen G., Saunders D., Sodergren E., Davis P.,
RA Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA Farbrother P., Desany B., Just E., Morio T., Roest R., Churcher C.,
RA Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,
RA Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,
RA Lousis H., Mungall K., Oliver K., Price C., Quail M.A.,
RA Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Sanders M.,
RA Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,
RA Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,
RA Shauleky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,
RA Chisholm R.L., Gibbs R., Loomis W.F., Platzner M., Kay R.R.,
RA Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;
RT "The genome of the social amoeba Dictyostelium discoideum.";
RL Nature 0:0-0(2005).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAF10100124; EAL64677.1; -; Genomic DNA.
RW Hypothetical protein.
SQ SEQUENCE 540 AA; 60799 MW; 0F4E391CB55A35B5 CRC64;

Query Match 13.6%; Score 102; DB 2; Length 540;
Best Local Similarity 29.4%; Pred. No. 25;
Matches 25; Conservative 18; Mismatches 38; Indels 4; Gaps 1;
```





```

RN RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC -!- INTERACTION:
CC O46085;EG:63B12.5; NbExp=1; IntAct=EBI-1233244, EBI-151469;
CC EMBL; AE003808; AAF58064.2; -; Genomic_DNA.
DR IntAct; Q9V719; -.
DR Ensembl; CG8421; Drosophila melanogaster.
DR FlyBase; FBgn0034075; Asph.
DR FlyBase; FBgn0034075; CG8421.
SQ SEQUENCE 556 AA; 63144 MW; B420980CBDC6357A CRC64;

Query Match 13.6%; Score 101.5; DB 2; Length 556;
Best Local Similarity 24.5%; Pred. No. 28;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

Qy 21 EDFILPVYKGELEKGYQFDGW-----EISGPEGKKDAGYVI-----NLSKDTFIK 65
Db 78 EDLDTPLSESPSK--VFDGWDEHRDHDGHDVQSPGSEALDDHDEHDDHDEDEDE 135
Qy 66 PVFKKIKBKEENKPT-----FDVSKKQNPQVNHSQLNESHKEDLQREHSQKSDS 119
Db 136 PLTSELEBELEBEERTEDEPAADSEYEEDEDEENNA--GNITAEADAEEBEEDND 193
Qy 120 TKDVTATVLDKNKISKST 138
Db 194 EGTVEATVEATTEATT 212

RESULT 35
Q7RLE7 PLAYO PRELIMINARY; PRT; 674 AA.
AC Q7RLE7;
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Notchless-related.
GN Name=PY02598;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Anguoli S.V., Suh B.B., Kooij T.W., Perteau M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabhi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABL01000712; EAA22065.1; -; Genomic_DNA.
DR HSSP; P16649; 1ERJ.
DR Pfam; IPR001680; WD40.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS00678; WD_REPEATS_2; 6.
DR PROSITE; PS00294; WD_REPEATS_REGION; 2.
KW Repeat; WD repeat.
SQ SEQUENCE 674 AA; 76535 MW; E7521B469F8E80F7 CRC64;
```

```

Query Match 13.6%; Score 101.5; DB 2; Length 674;
Best Local Similarity 23.1%; Pred. No. 34;
Matches 34; Conservative 27; Mismatches 51; Indels 35; Gaps 6;

Qy 9 NGKEMSGTIYSE-----EDFILPVYKGELEKGYQFDGWEIFSGPEGKKDAGYVINLSK 60
Db 500 NGKFIASSFDKSIIRIWSGIDGTLYAVFRGHVGPAYKI-AWSI-----DNNYIISCSQ 551
Qy 61 DTFIK-----PVFKKIBE-----KKEENKPTFDV--SKKQNPQVNHSQLNE 101
Db 552 DSTLKLWRINHLVPLAKKKGENGQPKSDPKDQKNDQKDDKDDKDDKDDKDDNDHANN 611
Qy 102 SHRKEDLQREHSQKSDSTKDVTATVL 128
Db 612 QEDGEEKKKKEKKEKDKIKSKIKITLL 638

RESULT 36
Q9GQ82 DROME PRELIMINARY; PRT; 785 AA.
AC Q9GQ82;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 10-MAY-2005 (TremBLrel. 30, Last annotation update)
DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA, isoform A).
GN Name=Asph; ORFNames=CG8421;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20564328; PubMed=10956665; DOI=10.1074/jbc.M006753200;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved
RT isoform of Asph missing the catalytic domain share exons with
RT junctin".
RL J. Biol. Chem. 275:39543-39554(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.M., Hoskins R.A., Galle R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Besen K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck H., Brockstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler J., Cantley E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos J.M., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fodor C., Giannelis A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
```

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426065; PubMed=12537568;  
RA Celnik S., Adams M., Champe M., Krommiller B., Carlson J.W., Halpern A.,  
RA Patel S., Adams M., Champe M., Dugan S.P., Friese E., Hodgson A.,  
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,  
RA Paclet J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
RA Svirekas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
RT melanogaster eukaryotic genome sequence.";  
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426070; PubMed=12537573;  
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J.W., Svirekas R.,  
RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
RA Ashburner M., Celnik S.E.,  
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
RT a genomics perspective.";  
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).  
RN [5]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=22426069; PubMed=12537572;  
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
RA Hradecky P., Huang Y., Millburn J.S., Millburn G.H., Prochman S.E.,  
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
RA Battencourt B.R., Celnik S.E., de Grey A.D.N.J., Drysdale R.A.,  
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
RA Lewis S.E.,  
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
RT systematic review.";  
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
RN [6]  
RP NUCLEOTIDE SEQUENCE.  
RG Berkeley *Drosophila* Genome Project;  
RA Celnik S., Carlson J., Wan K., Pfeiffer B., Friese E., George R.,  
RA Hoskins R., Stapleton M., Paclet J., Park S., Svirekas R., Smith E.,  
RA Yu C., Rubin G.,  
RT "Drosophila melanogaster release 4 sequence.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP NUCLEOTIDE SEQUENCE.  
RG FlyBase;  
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF289493; AAG40806.1; -; mRNA.  
DR EMBL; AE003808; AAM70947.1; -; Genomic DNA.  
DR Ensemble; CG8421; *Drosophila melanogaster*.  
DR FlyBase; FBgn0034075; Asph.  
DR FlyBase; FBgn0034075; CG8421.  
DR GO; GO:0030176; C:integral to endoplasmic reticulum membrane; IEA.  
DR GO; GO:0005488; F:binding; IEA.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0004597; P:peptide-aspartate beta-dioxygenase activity; IEA.  
DR GO; GO:0018193; P:peptidyl-amino acid modification; IEA.  
DR InterPro; IPR007803; Asp Arg Hydrol.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR InterPro; IPR001440; TPR-like\_helical.  
DR InterPro; IPR011990; TPR REGION; 1.  
DR Pfam; PF05118; Asp Arg Hydrol; 1.  
DR PROSITE; PS0293; TPR REGION; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.

SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD6836F7E1 CRC64;  
Query Match 13.6%; Score 101.5; DB 2; Length 785;  
Best Local Similarity 24.5%; Pred. No. 40;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
Qy 21 EDFTLPVYKGELEKGYQDGM-----EISGFEGKKDAGYVI-----NLSKDTFIK 65  
Db 78 EDLDTPLSERSPSK--VPDGMVDHRDHDGVDQEPSGEALDDHDDHDDHDEDEDEE 135  
Qy 66 PVFKKIBKKKEENKPT-----FDVSKKDNPNQVNSHRSKEDLQREHSQKSDS 119  
Db 136 PLTEEELEEESEETPEDEPAADSEYEDEDEENNA--GENITARDABEEEDND 193  
Qy 120 TKQVTVTLDKNNISSKST 138  
Db 194 EGTVEATVEATTEATTEAT 212  
RESULT 37  
Q6HNR0 BACHK  
ID Q6HNR0 BACHK PRELIMINARY; PRT; 954 AA.  
AC Q6HNR0;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Possible internalin protein.  
GN OrderedLocustNames=BR727\_0463;  
OS *Bacillus thuringiensis* (subsp. konkukian).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;  
OC *Bacillus cereus* group.  
OX NCBI\_TaxID=180856;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=97-27;  
RA Brettn T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
RA Richardson P., Rubin E., Tice H.,  
RT "Complete genome sequence of *Bacillus thuringiensis* 97-27.";  
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE017355; AAT63966.1; -; Genomic DNA.  
DR GO; GO:0009986; C:cell surface; IEA.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR001611; LRR.  
DR InterPro; IPR003885; LRR\_CYST.  
DR InterPro; IPR007092; LRR\_SDS22.  
DR InterPro; IPR006635; NEA\_transpt.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF00560; LRR\_1; 8.  
DR Pfam; PF05031; NEAT; 1.  
DR PRINTS; PR00019; LEURICHRPT.  
DR SMART; SM00365; LRR\_SD22; 8.  
DR SMART; SM00725; NEAT; 1.  
DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
DR TIGRFAMs; PS0978; NEAT; 1.  
DR Complete proteome.  
KW COMPLETE PROTEOME.  
SQ SEQUENCE 954 AA; 108589 MW; 4F4CF8B44C9B355F CRC64;  
Query Match 13.6%; Score 101.5; DB 2; Length 954;  
Best Local Similarity 25.7%; Pred. No. 50;  
Matches 43; Conservative 32; Mismatches 43; Indels 49; Gaps 9;  
Qy 21 EDFTLPVY--KGEL-----EKGYQP--DGWEIS-CFEG--- 48  
Db 727 EIIQVPVYDLEGEIENIKTSEDGTNNGVKNSTGPKVKYKFDLSDSEISFNGLVI 786  
Qy 49 -----KKDAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSKKDNPNQVNSH 100  
Db 787 QNIVEKEEKEPTKEVEESKEEKEPT-KEVEESKEEKEPTKEVEESKEEKEPTKEVE 845  
Qy 101 ESHKEDLQREHSQKSDSTKVTVTLVDKN-----NLSKSTTNPN 143  
Db 846 ES--KBEV--KEPTKEVEESKEEVAQBIKESKEEINQSPVQEQNVN 888



```

RESULT 38
Q9NFV9 PLAF9
ID Q9NFV9 PLAF9 PRELIMINARY; PRT; 329 AA.
AC Q9NFV9
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN Name=msp3;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=7G8;
RX MEDLINE=20416497; PubMed=10960178; DOI=10.1016/S0166-6851(00)00245-0;
RA Okeru D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252287; CAB85901.1; -; Genomic_DNA.
DR InterPro; IPR010784; Merozoite_SPAM.
DR Pfam; PF0133; Merozoite_SPAM; 1.
KW Merozoite.
FT NON TER 1
FT NON TER 329
SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;

Query Match 13.5%; Score 101; DB 2; Length 329;
Best Local Similarity 25.3%; Pred. No. 17;
Matches 38; Conservative 25; Mismatches 47; Indels 40; Gaps 7;

Qy 7 IQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGP--EGKKDAG-----YVINLS 59
Db 171 VLKAEAS-----YDYL-----GWFGGVPHEKKERMLSHLYVSSKD 211

Qy 60 KDTFIKPVFKIEKKBEENKPTFDVSKKONQVNHSQLNESHKRLQREH----- 113
Db 212 KENISKENDVDLDE-KKEEABETEELEKNEETSEISEDEEEKEEKEKEQAKE 270

Qy 114 --SQKSDTKDVTATVLKNNISSKSTNN 141
Db 271 QSNENNDDKXDMEA-----QNLISKNNNN 295

RESULT 39
Q5W5T1 TETPY
ID Q5W5T1 TETPY PRELIMINARY; PRT; 1015 AA.
AC Q5W5T1;
DT 01-FEB-2005 (TREMBlrel. 29, Created)
DT 01-FEB-2005 (TREMBlrel. 29, Last sequence update)
DT 01-FEB-2005 (TREMBlrel. 29, Last annotation update)
DE Putative P-glycoprotein (Fragment).
GN Name=mdr1;
OS Tetrahymena pyriformis.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymenidae; Tetrahymena.
OX NCBI_TaxID=5908;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GL;
RA Camares O., Denizau F., Bamdad M.;
RT "Characterisation of MDR sequence homologue in Tetrahymena pyriformis.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AJ514918; CAD5936.2; -; Genomic_DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0016887; F:ATPase activity; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. .; IEA.

```

```

DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR011527; ABC_membrane_1.
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transp_like.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS0929; ABC_TM1P; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS08993; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Nucleotide-binding.
FT NON TER 1
FT NON TER 1
SQ SEQUENCE 1015 AA; 114219 MW; CB2E9AB73768A778 CRC64;

Query Match 13.4%; Score 100.5; DB 2; Length 1015;
Best Local Similarity 28.3%; Pred. No. 63;
Matches 41; Conservative 24; Mismatches 57; Indels 23; Gaps 6;

Qy 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGWEISGFEGKKDAGYVINLSK 60
Db 317 HRLT-TIENADE-----ILVIDKGLVQGTFD--QLDARGKPELAKNQIQK 362

Qy 61 DTFIKPVFKIEKKBEENKPTFDVSKKONQVNHSQLNESH-RKEDLQREHSHKSDS 119
Db 363 EQKDEERRKKEQLQEDRK---DFEERMRKSTVGKSLIMENHLTKEQIAAEKEKEQAY 419

Qy 120 TKDVTATVLKNNISSKSTNNPNK 144
Db 420 FKE-----LDKNMVTRLFMTNRPER 439

RESULT 40
Q9PPL5 CAMJE
ID Q9PPL5 CAMJE PRELIMINARY; PRT; 312 AA.
AC Q9PPL5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Putative membrane protein.
GN OrderedLocusNames=Cj0692.3;
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204; DOI=10.1038/35001088;
RA Parkhill J., Wren B.W., Mungall K.L., Ketley J.M., Churcher C.M.,
RA Braham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagsels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139076; CAB72966.1; -; Genomic_DNA.
DR PIR; G81339; G81339.
KW Complete proteome.
SQ SEQUENCE 312 AA; 37221 MW; 0004FA7836A741B8 CRC64;

Query Match 13.4%; Score 100; DB 2; Length 312;
Best Local Similarity 25.3%; Pred. No. 19;
Matches 41; Conservative 32; Mismatches 61; Indels 28; Gaps 8;

Qy 7 IQNGKMSSTI---VSEEDFILPVYK-----GELEKGYQPDGWEISGFEGKKDAGYV 55
Db 24 INQKPLDDLDLRDEISSDDILRRPFKKTPNKFLDELDDEYEYKTKSNLYLKED---L 80

Qy 56 INL----SKDTFIKPVFKIEKKBEENKPT---FDVSKKONP-----QVNHSQLNESHK 105

```

Db 81 INVLEEKQSLAKIFSKMKERRKEENKTKKNPLFSRKKANEIKNIQTKTQIQTKSNQA 140  
Qy 106 EDLQREHSQKSDSTKV--TATVLDKNNISK--STTNPN 143  
Db 141 TTQTQEKKELTNSIEKIQTETKIQKPLIIEKKLDVKNQPN 182

Search completed: April 24, 2006, 14:59:35  
Job time : 88.4631 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 14:50:52 ; Search time 15.1409 Seconds  
(without alignments)  
915.083 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773  
Perfect score: 748  
Sequence: 1 HRVTVTIQNGKEMSTIVSE.....ATVLDKNNISSKSTNNPNK 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_80:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	748	100.0	2140	2 F95074	serine proteinase,
2	745	99.6	2144	2 A97942	metalloproteinase,
3	110	14.7	558	2 T18467	hypothetical prote
4	108.5	14.5	1038	2 JC5497	claustrin - chicke
5	104	13.9	665	2 B71609	hypothetical prote
6	103.5	13.8	325	2 T18283	hypothetical prote
7	100	13.4	312	2 G81339	probable membrane
8	99	13.2	211	2 T25911	hypothetical prote
9	97.5	13.0	1345	2 S46817	hypothetical prote
10	97	13.0	988	2 T14188	hypothetical prote
11	95.5	12.8	644	2 T47835	hypothetical prote
12	95	12.7	348	2 I37271	cylicin II - human
13	95	12.7	535	2 T37189	hypothetical prote
14	95	12.7	622	2 A90570	lipoprotein (impor
15	95	12.7	1397	2 T10466	DNA topoisomerase
16	94.5	12.6	3724	2 T18427	hypothetical prote
17	94	12.6	210	2 T28771	hypothetical prote
18	92.5	12.4	219	2 B72291	hypothetical prote
19	92.5	12.4	540	2 D86432	hypothetical prote
20	91	12.2	253	2 T32879	hypothetical prote
21	91	12.2	385	2 T20410	hypothetical prote
22	90.5	12.1	614	2 A84152	hypothetical prote
23	90.5	12.1	670	2 T28391	ORF MSV230 hypothe
24	90.5	12.1	1332	2 S41552	probable transcrip
25	90.5	12.1	2401	2 T28676	rhopty protein -
26	89.5	12.0	508	2 T81594	hypothetical prote
27	89.5	12.0	508	2 E86549	hypothetical prote
28	89.5	12.0	508	2 T27074	hypothetical prote
29	89.5	12.0	700	2 S67610	probable membrane

30	89.5	12.0	1016	2 T19006	ankyrin related pr
31	89	11.9	528	2 E96795	unknown protein F2
32	89	11.9	645	2 E89883	conserved hypothet
33	89	11.9	1888	2 T39009	hypothetical prote
34	88.5	11.8	301	2 T33068	hypothetical prote
35	88.5	11.8	371	2 A71683	hypothetical prote
36	88.5	11.8	762	2 G88436	protein T04A8.13 [
37	88.5	11.8	791	2 T24435	hypothetical prote
38	88	11.8	276	2 T23451	hypothetical prote
39	88	11.8	1280	2 G96796	hypothetical prote
40	87.5	11.7	1702	2 A41859	IGA-specific metal
41	87.5	11.7	2523	2 T18477	hypothetical prote
42	87	11.6	635	2 F71621	hypothetical prote
43	87	11.6	2464	1 QKMSB1	microtubule-associ
44	87	11.6	2708	2 T09079	probable chloroqui
45	86.5	11.6	792	2 T49989	hypothetical prote

ALIGNMENTS

RESULT 1

F95074  
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: F95074  
R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; He  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
naon, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: F95074  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2140 <KUR>  
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AB005672; PIDN:AAK74791.1  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP0641

Query Match	100.0%;	Score	748;	DB	2;	Length	2140;
Best Local Similarity	100.0%;	Pred. No.	9.4e-49;				
Matches	144;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	HRVTVTIQNGKEMSTIVSE	DFILPVYKGELEKGYQFDGWEISGPEKGDAGYVNL	SK	60		
Db	1963	HRVTVTIQNGKEMSTIVSE	DFILPVYKGELEKGYQFDGWEISGPEKGDAGYVNL	SK	2022		
Qy	61	DTFTKPVKKEEKKENKPT	FDVSKKQNPQVNHSQLNESHKEDLQREHSQKSD	ST	120		
Db	2023	DTFTKPVKKEEKKENKPT	FDVSKKQNPQVNHSQLNESHKEDLQREHSQKSD	ST	2082		
Qy	121	KDVTATVLDKNNISSKSTNNPNK	144				
Db	2083	KDVTATVLDKNNISSKSTNNPNK	2106				

RESULT 2

A97942  
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C:Accession: A97942  
R:Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.;  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.;  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: A97942  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2144 <KUR>  
A;Cross-references: UNIPROT:Q8DQP7; UNIPARC:UPI00000E3490; GB:AE007317; PIDN:AAK99365.1;  
C;Genetics:  
A;Gene: prtA  
C;Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 745; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 1.6e-48;  
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 HRVTIIONGKEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 60  
Db 1967 HRVTIIONGKEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 2026  
  
Qy 61 DTFIKPVPKKIEEKKKEENKPTFDVSKKQNPQVNHSQLNESHKREDIQRHEHSQKSDST 120  
Db 2027 DTFIKPVPKKIEEKKKEENKPTFDVSKKQNPQVNHSQLNESHKREDIQRHEHSQKSDST 2086  
  
Qy 121 KDVTATVLDKNNISSKSTNNPNK 144  
Db 2087 KDVTATVLDKNNISSKSTNNPNK 2110

RESULT 3  
T18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18467  
R;Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A;Reference number: Z18937  
A;Accession: T18467  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-558 <LAW>  
A;Cross-references: UNIPROT:077355; UNIPARC:UPI000017CC2E; EMBL:AL008970; NID:e1407852;  
C;Genetics:  
A;Map position: 3  
A;Introns: 84/1; 160/1  
A;Note: C0465c

Query Match 14.7%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 0.54;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;  
  
Qy 17 IVSEEDFILPVY----KGELEKGYQFDGWEISGFEGKK---DAGYVINLSKDTFIKPV 67  
Db 60 IIGFEDDILYECISQLQSKSEK--KADGEDKYLNAKKLKLNTLGTGFGNKKSDIFIEL 117  
  
Qy 68 FKKI--BEKGE-----ENKPTDVSQ-KKQNPQVNHSQLN-----SHRK 105  
Db 118 LELLNEEKGEHIADTLNENK-TWDIKKKNENINENYVNEKQISNKDKHVSQHN 176  
  
Qy 106 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTTN 140  
Db 177 EHNINNVNLKKEKYTDIQDKRKHRSLSQSDSYKKRPFNKRKTSTIER-SLSNKRDE 235  
  
Qy 141 NPNK 144  
Db 236 KTNK 239

RESULT 4  
JC5497  
claustrin - chicken  
N;Alternate names: keratan sulfate proteoglycan  
C;Species: Gallus gallus (chicken)  
C;Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 09-Jul-2004  
C;Accession: JC5497; PC4334; S37561

R;Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A;Title: Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A;Reference number: JC5497; MUID:94157526; PMID:7906711  
A;Accession: JC5497  
A;Molecule type: mRNA  
A;Residues: 1-1038 <BURI>  
A;Cross-references: UNIPROT:Q90784; UNIPARC:UPI00000PC87F; EMBL:X67778; NID:9406318; PID:  
A;Accession: PC4334  
A;Molecule type: protein  
A;Residues: 79-83;298-412;485-502 <BUR2>  
A;Cross-references: UNIPARC:UPI000017BFF3; UNIPARC:UPI000017BFF4; UNIPARC:UPI000017BFF5  
A;Experimental source: brain  
C;Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervo  
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F;267-270/Region: cell attachment (R-G-D) motif  
F;112,213,490/Binding site: carbohydrates (Asn) (covalent) #status predicted  
F;152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predict

Query Match 14.5%; Score 108.5; DB 2; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 1.4;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;  
  
Qy 11 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVK- 69  
Db 595 KPETKTIVAEKDV-----TTKEKQLGKSETSEKQSEKQDVKPKVTKESVKCEVKA 646  
  
Qy 70 KIEBKKEENKPTFDVSKKQNPQVNHSQLNESHKREDIQR-----EHSQKSDSTKDV 123  
Db 647 KPEEKKDEKPKKKEVSKKEEKLPI---KKBEKPKQSDIKKEVKCKVKEEKKAKKEV 702

RESULT 5  
B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 09-Jul-2004  
C;Accession: B71609  
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
.; Fertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A;Reference number: A71600; MUID:99021743; PMID:9804551  
A;Accession: B71609  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-665 <GAR>  
A;Cross-references: UNIPROT:O96229; UNIPARC:UPI000017B60A; GB:AE001410; GB:AE001362; NID:  
A;Experimental source: clone 3D7  
C;Genetics:  
A;Gene: PFB0680w

Query Match 13.9%; Score 104; DB 2; Length 665;  
Best Local Similarity 25.9%; Pred. No. 1.9;  
Matches 42; Conservative 29; Mismatches 35; Indels 56; Gaps 10;  
  
Qy 1 HRVT-VTIQNKEMSTIVSEDFILPVYKGELEKGYQFDGWEI--SGFEGKKDAGYVIN 57  
Db 138 HRQNELNLQSGK-----NEQDI-----NKNEKGKQ----DISNSNAENKKD----- 174  
  
Qy 58 LSKDTFIKPVFKKIEEKK-----BENKPTFD-----VSKKQNPQVNHSQLNESH 103  
Db 175 -----VKEGVKELEEKKEEKISDD:HKVEENKSDDKHVEENKSDDKHVEENKSDDH 228  
  
Qy 104 RKEDLQR-EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 144  
Db 229 KIEEVKKVEEHEDEER-----DKGKKSENKKNKDNK 261

RESULT 6  
T18283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C;Species: Dictyostelium discoideum

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18283  
R;Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh  
Genetics 148, 1117-1125, 1998  
A;Title: Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to the Ddp1  
A;Reference number: Z14684; MUID:98198836; PMID:9539429  
A;Accession: T18283  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-325 <RIB>  
A;Cross-references: UNIPROT:O44016; UNIPARC:UPI000007C824; EMBL:U00796; NID:g2702254; PI  
C;Genetics:  
A;Introns: 85/1

Query Match 13.8%; Score 103.5; DB 2; Length 325;  
Best Local Similarity 23.8%; Pred. No. 0.91;  
Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

QY 1 HRVTVTIQKEMSTIVSEEDFILPVYK-GEL--EKGYQDGEWISGFEKG----- 49  
DB 57 HRTITSIKN--RFSVKKIGDEEKLFRISKNGELIVINLELFDNPHIK--EGKHLRKSQMF 112  
QY 50 ---KDAGYV-----INLSKDTFTKPV-----FKK----- 70  
DB 113 NHIKDSGYATNVEIEIFLESCITLCKBITAQTAKNSYKRNIIINKLPREHEHEHEHEHE 172  
QY 71 --IEEKEENKPTPDVSKKONPOVNSQLNESHKE-----DLQREHSQKSDSTK 121  
DB 173 EEEERQEEVEKPTISSEEBETPAVSEEEKEEBEEETPAVSEEEKEEBEQEDKEK 232  
QY 122 D 122  
DB 233 D 233

RESULT 7  
G81339  
probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)  
C;Species: Campylobacter jejuni  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C;Accession: G81339  
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell  
Nature 403, 665-668, 2000  
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A;Reference number: A81250; MUID:20150912; PMID:10688204  
A;Accession: G81339  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-312 <PAR>  
A;Cross-references: UNIPROT:Q9PPL5; UNIPARC:UPI00000C1CF0; GB:AL111168; NID  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj0692c

Query Match 13.4%; Score 100; DB 2; Length 312;  
Best Local Similarity 25.3%; Pred. No. 1.6;  
Matches 41; Conservative 32; Mismatches 61; Indels 28; Gaps 8;

QY 7 IQNGKEMSTI---VSEEDFILPVYK-----GELEKGYQDGEWISGFEKGKDGAYV 55  
DB 24 INQKPLDDDLRDEISSDILRRFPKKTTPNKFLEELDEYESKHTKSNLYLKED---L 80  
QY 56 INL---SKDTFTKPVFKIEEKEENKPT---PDVSKKONP---OVNHSQKSNHRK 105  
DB 81 INVLEEKQSLAKLIFSGKKERKEENKKTKNFLFSRKKANEIKNIQTQIQTKSNQA 140  
QY 106 EDLQREHSQKSDSTKDV--TATVLDKNNISSK--STTNPN 143  
DB 141 TTQTKQKKELTNSIEKIQKTETKIQKPLIEKKLDVKNQPN 182

RESULT 8

## T25911

hypothetical protein T23B3.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T25911  
R;Maggi, L.; Le, T.  
submitted to the EMBL Data Library, February 1997  
A;Description: The sequence of C. elegans cosmid T23B3.  
A;Reference number: Z20109  
A;Accession: T25911  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-211 <MAG>  
A;Cross-references: UNIPROT:P91488; UNIPARC:UPI000007D762; EMBL:U88309; PIDN:AAB42334.1  
A;Experimental source: strain Bristol N2; clone T23B3  
C;Genetics:  
A;Gene: CESP:T23B3.5  
A;Map position: 1  
A;Introns: 30/2; 200/3

Query Match 13.2%; Score 99; DB 2; Length 211;  
Best Local Similarity 29.4%; Pred. No. 1.2;  
Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

QY 47 EGKQDAGYVINLSKDTFTKPVFKIREEKEENKPTFDVSKKONPOVNH-----SQLNES 102  
DB 69 EGEKKQGEKKSEKQDGKGEKEKQDGKQDQKQDQKQDQKQDQKQDQKQDQKQDQK 128  
QY 103 HRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 144  
DB 129 EKQDKQK 170

## RESULT 9

S46817  
hypothetical protein YHR080c - yeast (Saccharomyces cerevisiae)  
C;Species: Saccharomyces cerevisiae  
C;Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 05-Oct-2004  
C;Accession: S46817  
R;Favella, T.  
submitted to the EMBL Data Library, June 1994  
A;Description: The sequence of S. cerevisiae cosmid 9205.  
A;Reference number: S46795  
A;Accession: S46817  
A;Molecule type: DNA  
A;Residues: 1-1345 <FAV>  
A;Cross-references: UNIPROT:P38800; UNIPARC:UPI000013B2B1; EMBL:U10556; NID:g500825; PI  
C;Genetics:  
A;Cross-references: SGD:S0001122  
A;Map position: 8R  
C;Superfamily: uncharacterized conserved protein  
C;Keywords: transmembrane protein

Query Match 13.0%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 13;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 29 KGELEKGYQDGEWISGFEKGK-DAGYVINLSKDTFTKPVFKIEEKEENKPTFDVSK 87  
DB 1109 KGATEKG-----SVGEQKVSVDVNLSELRII-----SRKSKKPVKKYWK 1149  
QY 88 KCONPOVNSQLNESHKEKDLQREHSQKSDSTKDVATVLD--KNNISSKSTTN---NP 142  
DB 1150 SHDKHRFPFSKVE-----QKSSEKSKSDDNKILTHILDPOVNNFSSBIFNKLISP 1201  
QY 143 NK 144  
DB 1202 QK 1203

## RESULT 10

T14188  
hypothetical protein T28D5.30 - Arabidopsis thaliana



C;Species: *mezobrya maritima*  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C;Accession: B72291  
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hicke

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence  
A;Reference number: A72200; MUID:99287316; PMID:10360571  
A;Accession: B72291  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-219 <ARN>  
A;Cross-references: UNIPROT:Q9X0M6; UNIPARC:UPI000000C12ED; GB:AE001771; GB:AE000512; NID  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TW1142

Query Match	12.4%	Score 92.5;	DB 2; Length 219;
Best Local Similarity	28.1%;	Pred. No. 4;	
Matches 36; Conservative 24; Mismatches 35; Indels 33; Gaps 8;			
Qy	30	GELEKGQV--FDGWEISG-----PEGKDGAGYVIN-LSKDFTFKPVPFKEEKKRENK	80
Dd	89	GEEVSVRDLPAGEGVGRGTPTFPFFPKGEGUGLYPGYYVDKNDFIK-IILKYVAQLKEED--	145
Qy	81	PTFDVSKKKONPVNHSQNESHRED---LQREHSSQSDDS-----TKD	122
Dd	146	--FQTALVKODPFVEGPLIITIF-KEDADVLFXDENAVKVDTYPNVRRDRIVYTDSPD	202
Qy	123	VTAATVLDK	130
Dd	203	VAKTLOBK	210

RESULT 19  
D86432  
hypothetical protein TS18.14 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: D86432  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
anssen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: D86432  
A:Status: preliminary  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-540 <STO>  
A:Cross-references: UNIPROT:Q9SA84; UNIPARC:UPI000000ABF35; GB:A0505172; NID:94587525; P  
C:Genetics:  
A:Map position: 1

	Query Match	12.4%	Score 92.5	DB 2	Length 540
	Best Local Similarity	24.8%	Pred. No. 11		
	Matches 32	Conservative 25	Mismatches 55	Indels 17	Gaps 4
Qy	10	GKEMSSITIVSEDFILPVYKGELEKGYQFDGWEISGPEGKDGAVINLSKDTFIKPVFK	69		
Db	47	GVEKSASFKESDFPADLKESEK-----AUSDLSKLEEAIVN-----TLTKTKKK	94		
Qy	70	K---IESKKEENKPTFVDSKKDKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTAT	126		
Db	95	ESSPMKEKGEVVRPEAEVEKKKE--EAAEKVTEEEKSEAVVTEEPKAEVTEAVVTEE	152		
Qy	127	VLDKNNISS	135		
Db	153	IIPKEEVT	161		

```

RESULT 20
T32879
hypothetical protein C17F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T32879
R:Gattung, S.; Scheet, P.
A:Submitted to the EMBL Data Library, January 1998
A:Description: The sequence of C. elegans cosmid C17F3.
A:Reference number: Z21240
A:Accession: T32879
A:Status: preliminary; translated from 3B/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-253 <GAT>
A:Cross-references: UNIPROT:O44948; UNIPARC:UPT0000074BB9; EMBL:AF043692; PIDN:
A:Experimental source: strain Bristol N2; clone C17F3
C:Genetics:
A:Gene: CESP:C17F3.3
A:Map position: 1
A:Introns: 41/1

Query Match          12.2%;   Score 91;   DB 2;   Length 253;
Best Local Similarity 32.6%;   Pred. No. 6.1;   33;   Indels 24;   Gaps 3;
Matches 31;   Conservative

Qy 66  PVFKKTEEKKE---EENKPTFDVSKK-----KONPQVNHSQLNESHK---- 104
Db 61  PVAPKVEEKKEKKEKKADDEKKITEEKDDKKSKKTEEKDKKISVKKTQETKSKERDKK 120

Qy 105 ----KEDLQREHSQKSDSTKQDVATVTLDKNNISSK 136
Db 121 DEKDEKKEESKEKSKDEKKDKKDKYDKKKEDEKK 155

```

```

RESULT 21
T20410
hypothetical protein E02A10.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20410
R:Thomas, K.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19271
A:Accession: T20410
A:Molecule type: DNA
A:Status: preliminary; translated from c3B/EMBL/DDBJ
A:Residues: 1-385 <MIL>
A:Cross-references: UNIPROT:Q93424; UNIPARC:UPI00000835C8; EMBL:Z810S3; PIDN:CAI
A:Experimental source: clone E02A10
C:Genetics:
A:Gene: CESP:E02A10.2
A:Map position: 5
A:Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 12.2%; Score 91; DB 2; Length 385;
Best Local Similarity 29.9%; Pred. No. 9.8;
Matches 38; Conservative 22; Mismatches 41; Indels 26; Gaps 6;

Qy 4 TVTQNGKMSSTVSEEDFILPVYKGELEKYGQFGWISGFGKKDAGYVNLKDTF 63
|: | | | | | | | | | | | | | | | | | | | |
Db 271 TILRINHKEVAEK--NEEDKKEPKEEKKKEVEKKE---EDEKKDE----- 314

Qy 64 IKPVFKIEKKKEENKPTFVSKKNDNPQVNHSQLNESHRRKEDLQREHHSQKSDSTKDV 123
|: | | | | | | | | | | | | | | | | | | | |
Db 315 -EP--KKEEKKEEQQE--EVEKKE-----BEKKDEEPKKEEKKKEEKKDEVEEK 364

Qy 124 TATVLDK 130
|: | | |
Db 365 SEKVEEK 371
|: | | |

RESULT 22
A84152

```





```
Db 549 ITRNRADLEKEI-----EDM-----EKDKDYELDEBEVAGSRKG-----LNMGAHMLA 594
Qy 65 K--VFVKKIEEKKEENKPTFD-----VSKKDD-----NPQVNHSQLNESH 103
Db 595 KENGKVSSEKSKTKVDRAPTNDKLTSPVPEGEKEKDKTASSVTTVHENVNKNKEIKENG 654
Qy 104 RKEDLQR-EESHSQKSDSTKDVATVLD-KNNISSKSTTN 140
Db 655 KNEEQDMVEESKTESKDAADAAKQDYGLODKTAEN 693

RESULT 25
T28676
rhoptry protein - Plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T28676; A45521
R:Singha, K.A.; Keen, J.K.; Ogun, S.A.; Hojder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507; MUID:97077455; PMID:8920022
A:Accession: T28676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:G1041784; PI
R:Keen, J.; Hojder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KEE>
A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

Query Match 12.1%; Score 90.5; DB 2; Length 2401;
Best Local Similarity 27.4%; Pred. No. 82;
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;

Qy 6 TQNGKMSSTI-----VSDEDILPVYKELEKGY---QFDGWEISGFEKKDA 52
Db 924 TSKNHEEKISDIRKNSLKIIDFSESRYINDI-KKELEKNVLESNNNTDINQVLSKIEN 982
Qy 53 GVINLSKDTFTKPVFKIEEKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQRE 112
Db 983 IY--NILKLNKIKKIIDKVKYETDIEK-----NNKKNVLESSEKIITQKENSLSKE 1035
Qy 113 HSQKSDSTKD-----VTATVLDKNNISS 135
Db 1036 COSKIKSTIDNNYVSECINKNTNLKTYIVNEKNNT 1072

RESULT 26
B81594
hypothetical protein CP0281 [imported] - Chlamydia pneumoniae (strain AR39)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: B81594
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: B81500; MUID:20150255; PMID:10684935
A:Accession: B81594
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <REA>
A:Cross-references: UNIPROT:Q9JRY3; UNIPARC:UPI00000D2FA8; GB:AE002189; GB:AE002161; NID
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0281
```

```
Query Match 12.0%; Score 89.5; DB 2; Length 508;
Best Local Similarity 24.5%; Pred. No. 17;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 64 IKPVFKKIEEKKEENKPTFD-----VSKKKNPQVNHSQLNESHK 105
Db 92 VKGVFKTTPQARPEVSSPRLPSHVQHQLPGLGEGFDRDRIQKRSENPEADLGKMKRSYSD 151
Qy 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139
Db 152 GDLDVRVGHDSNEDSTEDSRS---EGGEPSPSKSS 182

RESULT 27
B86549
hypothetical protein CP10473 [imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: B86549
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ie
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: B86549
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <STO>
A:Cross-references: UNIPROT:Q9JRY3; UNIPARC:UPI00000D2FA8; GB:BA000008; NID:G8978843; PI
A:Experimental source: strain J138
C:Genetics:
A:Gene: CP10473

Query Match 12.0%; Score 89.5; DB 2; Length 508;
Best Local Similarity 24.5%; Pred. No. 17;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 64 IKPVFKKIEEKKEENKPTFD-----VSKKKNPQVNHSQLNESHK 105
Db 92 VKGVFKTTPQARPEVSSPRLPSHVQHQLPGLGEGFDRDRIQKRSENPEADLGKMKRSYSD 151
Qy 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139
Db 152 GDLDVRVGHDSNEDSTEDSRS---EGGEPSPSKSS 182

RESULT 28
C72074
hypothetical protein - Chlamydia pneumoniae (strain CWL029)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: C72074
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: C72074
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <ARN>
A:Cross-references: UNIPROT:Q92878; UNIPARC:UPI00000C11CC; GB:AE001632; GB:AE001363; NID
A:Experimental source: strain CWL029
C:Genetics:
A:Gene: CPn0473

Query Match 12.0%; Score 89.5; DB 2; Length 508;
Best Local Similarity 24.5%; Pred. No. 17;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 64 IKPVFKKIEEKKEENKPTFD-----VSKKKNPQVNHSQLNESHK 105
Db 92 VKGVFKTTPQARPEVSSPRLPSHVQHQLPGLGEGFDRDRIQKRSENPEADLGKMKRSYSD 151
Qy 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139
```



A:Accession: E89883  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-645 <KUR>  
A:Cross-references: UNIPROT:Q99UX5; UNIPARC:UPI00000CAAE3; GB:BA000018; PID:g13700929; E89883  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0976

Query Match 11.9%; Score 89; DB 2; Length 645;  
Best Local Similarity 21.9%; Pred. No. 25;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;  
  
QY 6 TIONCKMSSTIVSEDFILPVYKGLB-KGVQF-----DGWEISGPEGKK-----50  
DB 355 SVNNESMMDTFVKH-----PIKTGLNGKRYVMVMTNDDYWKDFMVEGQRVTRISKDA 409  
QY 51 -----DAGVIVNL-SKDTPIKVFVKKIEKKEE 78  
DB 410 KNNRTIIPVYEGKTLDAIVKVHVHTIDYDQYHVRIVDKAFTKANTDKSNKKEOOD 469  
QY 79 NKPTFDV-----SKKDNQPNVHNSQLNESHKREDIQ-----REHSQKSDSTKQVT-ATVL 128  
DB 470 NSAKKEATPATPSKPTPSPEKESQKQSDKNQKLPSEKENDASSSESGDKTPTATKP 529  
QY 129 DKNNTSSKSTTNPNK 144  
DB 530 TKGVESSSTT--PTK 543

RESULT 33  
T39009  
hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
A:Accession: T39009  
R:Gentles, S.; Churcher, C.M.; Barrall, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: Z21815  
A:Accession: T39009  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1888 <GEN>  
A:Cross-references: UNIPROT:O14207; UNIPARC:UPI000013AR2F; EMBL:Z98531; PIDN:CAB11064.1;  
A:Experimental source: strain 972h-; cosmid c6B12  
C:Genetics:  
A:Gene: SPDB:SPAC6B12.02c  
A:Map position: 1  
A:Superfamily: Schizosaccharomyces hypothetical protein SPAC6B12.02c

Query Match 11.9%; Score 89; DB 2; Length 1888;  
Best Local Similarity 23.9%; Pred. No. 82;  
Matches 37; Conservative 29; Mismatches 59; Indels 30; Gaps 7;  
  
QY 14 SSTIVSEEDF-----ILPVY---KGELEKGYQPDGWEISGFE-----GKDAQ 53  
DB 390 SSSLTSENPFQLVNAAVSTIPVTRTYTKWKK-RFYVEVEKLPDLILLESYGKKAPK 448  
QY 54 YVINLSKDTFIKVPFKKIEKKEENKPTFDVSKKDNQPNVHNSQLNESHKREDIQREH 113  
DB 449 FLRVFARSSSHIP--KMIRRKQMDSKKYFSPDKESDRQVIDQVLSDWYSGKHELVOQSH 506  
QY 114 SOKSDS-TKQVATVLDKN-----NISKSTTN 141  
DB 507 SYKKSPDSKSGVGNIFSVNSKKHSHVINAATAANN 541

RESULT 34  
T33068  
hypothetical protein C35E7.9 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
A:Accession: T33068

R:Graves, T.; McDonald, R.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of C. elegans cosmid C35E7.  
A:Reference number: Z21278  
A:Accession: T33068  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-301 <GRA>  
A:Cross-references: UNIPROT:O61765; UNIPARC:UPI0000074979; EMBL:AF067216; PIDN:AAC17524.  
A:Experimental source: strain Bristol N2; clone C35E7  
C:Genetics:  
A:Gene: CESP:C35E7.9  
A:Map position: 1  
A:Introns: 30/3; 193/1; 236/2

Query Match 11.8%; Score 88.5; DB 2; Length 301;  
Best Local Similarity 24.8%; Pred. No. 12;  
Matches 27; Conservative 17; Mismatches 38; Indels 27; Gaps 3;  
  
QY 35 GYQFDGWEISGPEGKK-----DAGVIVNL-SKDTFIKVPFKKI-----EE 73  
DB 10 GITTAGILAGCGGKKKKGKSTASAAAPKADSKMPPVENVKSKKSEKKEPKKEEP 69  
QY 74 KKEENKPTFDVSKKDNQPNVHNSQLNESHKREDIQREHSQKSDSTKD 122  
DB 70 KKEEKKKSKSEKDKDK-----EEAKKEDDKDEKDEKDKD 112

RESULT 35  
A71683  
hypothetical protein RP278 - Rickettsia prowazekii  
C:Species: Rickettsia prowazekii  
C:Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 09-Jul-2004  
A:Accession: A71683  
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alemark, U.  
Nature 396, 133-140, 1998  
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.  
A:Reference number: A71630; MUID:99039499; PMID:9823893  
A:Accession: A71683  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-371 <AND>  
A:Cross-references: UNIPROT:Q9ZDP9; UNIPARC:UPI000000C1161; GB:AJ235269; NID  
A:Experimental source: strain Madrid E  
C:Genetics:  
A:Gene: RP278

Query Match 11.8%; Score 88.5; DB 2; Length 371;  
Best Local Similarity 23.1%; Pred. No. 15;  
Matches 34; Conservative 33; Mismatches 37; Indels 43; Gaps 9;  
  
QY 11 KEMSTTVSDEDFILPVYKGELEKGYQPDGWEISGFEKK-DAGVIVNL-SKDTFIKVPVK 69  
DB 192 EQLNKTITNEE-----FRKMLEQ-----EKKKIEGKKTQASEALNKK-----LKPIYK 234  
QY 70 KIEEKEE-----ENKPTFDVSKKKNPVN-----HSOLNESHKREDIQREHSQKSDSTK 121  
DB 235 QMDEBELFKLAILFOY-----AQANIDKHALYAKQYQTKIENDPNYKELEKIQ 286  
QY 122 DVTATVLDKNNISSKSTT-----NNFN 143  
DB 287 E-----IVSKIEYSKSKNTNDIINNPN 309

RESULT 36  
G88436  
protein T04A8.13 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
A:Accession: G88436  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog

A:Reference number: A75000; MUID:99069613; PMID:9851916  
A>Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: G88436  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-762 <STO>  
A:Cross-references: UNIPROT:Q22142; UNIPARC:UPI000017A5C4; GB:chr\_III; PIDN:CAA84732.1;  
C:Genetics:  
A:Gene: T04A8.13  
A:Map position: 3

Query Match 11.8%; Score 88.5; DB 2; Length 762;  
Best Local Similarity 25.2%; Pred. No. 33;  
Matches 32; Conservative 25; Mismatches 65; Indels 5; Gaps 3;  
QY 19 SEEDPILPVYKGELEKGYQFGWEISGPEGK-KDAGYVINLSKDTFIPKVPKKEEKEE 77  
DB 14 SGKQIMKFGY--DKKEGLGMQKEIVGDDKKDKARKERKIQDEFAB--LKQDEEKKKE 69  
QY 78 ENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDKNNISSKS 137  
DB 70 EAEKKEKKEKKEEKKEDGHEKKEDKKEDKKENENDEKSKSDKKESKEDKKEKTK 129  
QY 138 TTNNPNK 144  
DB 130 TEDNEGK 136

RESULT 37  
T23435  
hypothetical protein T04A8.13 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T24435  
R:Palmer, S.  
submitted to the EMBL Data Library, August 1994  
A:Reference number: Z19889  
A:Accession: T24435  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-791 <WIL>  
A:Cross-references: UNIPROT:Q22142; UNIPARC:UPI0000061132; EMBL:Z35663; PIDN:CAA84732.2;  
A:Experimental source: clone T04A8  
C:Genetics:  
A:Gene: CESP:T04A8.13  
A:Map position: 3  
A:Introns: 31/3; 212/1; 229/3; 331/3; 406/1; 472/3; 572/1; 651/2

Query Match 11.8%; Score 88.5; DB 2; Length 791;  
Best Local Similarity 25.2%; Pred. No. 34;  
Matches 32; Conservative 25; Mismatches 65; Indels 5; Gaps 3;  
QY 19 SEEDPILPVYKGELEKGYQFGWEISGPEGK-KDAGYVINLSKDTFIPKVPKKEEKEE 77  
DB 14 SGKQIMKFGY--DKKEGLGMQKEIVGDDKKDKARKERKIQDEFAB--LKQDEEKKKE 69  
QY 78 ENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDKNNISSKS 137  
DB 70 EAEKKEKKEKKEEKKEDGHEKKEDKKEDKKENENDEKSKSDKKESKEDKKEKTK 129  
QY 138 TTNNPNK 144  
DB 130 TEDNEGK 136

RESULT 38  
T23451  
hypothetical protein K08B3.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T23451  
R:McMurray, A.

submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19743  
A:Accession: T23451  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-276 <WIL>  
A:Cross-references: UNIPROT:Q9KUT1; UNIPARC:UPI0000077D88; EMBL:Z81568; PIDN:CAB04590.1  
C:Genetics:  
A:Gene: CESP:K08B3.2  
A:Map position: 3  
A:Introns: 81/3; 102/3; 169/1; 211/2

Query Match 11.8%; Score 88; DB 2; Length 276;  
Best Local Similarity 31.5%; Pred. No. 11;  
Matches 28; Conservative 10; Mismatches 29; Indels 22; Gaps 3;  
QY 58 LSKDTFIPKVPKKEEKEE--NKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQ 115  
DB 60 LEKDEDVKKKCKGKEEKEEKKEDKKEEKDD-----SKEDKKDEDHEK 107  
QY 116 KSDSTKDVATVLDKNNISSKSTTNNPNK 144  
DB 108 KK-----TAEKKEKKEKDKENKKNK 128

RESULT 39  
G96796  
hypothetical protein P28016.18 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: G96796  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, K.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G96796  
A:Molecule type: DNA  
A:Residues: 1-1280 <STO>  
A:Cross-references: UNIPROT:Q9SRD2; UNIPARC:UPI000000A3260; GB:AE005173; NID:G6143896; F:  
C:Genetics:  
A:Gene: F28016.18  
A:Map position: 1

Query Match 11.8%; Score 88; DB 2; Length 1280;  
Best Local Similarity 23.7%; Pred. No. 63;  
Matches 32; Conservative 24; Mismatches 61; Indels 18; Gaps 3;  
QY 8 QNGKEMSTTVSEEDPILPVYKGELEKGYQFGWEISGPEGKDGAGYVINLSKDTFIPKVP 67  
DB 110 KGGGSVFALLDDDD-----EKEDNESDGDCKDDEPVSFTGKKJASKKKGGKGSFAASA 164  
QY 68 F-----KKTEEKKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKS 117  
DB 165 FDALGSDDDDTTEVEHDEEESPTTSGKKKSS---KSSKKNNTNSFTADLLDEEGTDA 221  
QY 118 DSTKDVATVLDKNN 132  
DB 222 SNRSDENTTEDEES 236

RESULT 40  
A41859  
19A-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae  
C:Species: Haemophilus influenzae



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:36:22 ; Search time 11.2752 Seconds  
(without alignments)  
561.980 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTIIONGKMSSTIVSE.....ATVLDKNNISSTNNPNK 144

Scoring table: BLASUM62

Gapop 10.0 , Gapext 0.5

Searched: 225428 segs, 44002918 residues

Total number of hits satisfying chosen parameters: 225428

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA\_New.\*  
1: /SIDSS/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /SIDSS/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /SIDSS/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /SIDSS/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
5: /SIDSS/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
6: /SIDSS/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /SIDSS/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /SIDSS/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	14.7	746	6	US-10-793-626-652
2	108	14.4	354	7	US-11-189-817-2
3	105	14.0	651	7	US-11-128-660-1
4	89	11.9	501	6	US-10-485-517-381
5	89	11.9	645	6	US-10-485-517-244
6	87	11.6	140	7	US-11-096-568A-4771
7	86	11.5	443	7	US-11-188-298-1015
8	86	11.5	700	7	US-11-196-475-74
9	85.5	11.4	886	7	US-11-087-099-11456
10	85	11.4	472	6	US-10-793-626-658
11	85	11.4	700	7	US-11-196-475-66
12	85	11.4	708	7	US-11-196-475-76
13	85	11.4	720	6	US-10-793-626-2058
14	84.5	11.3	1694	7	US-11-052-554A-83
15	83.5	11.2	493	7	US-11-096-568A-3070
16	83.5	11.2	493	7	US-11-096-568A-3071
17	83.5	11.2	510	7	US-11-096-568A-3069
18	82	11.0	439	7	US-11-188-298-15964
19	82	11.0	439	7	US-11-188-298-15964
20	81.5	10.9	8746	7	US-11-098-686-10232
21	81	10.8	785	6	US-10-793-626-264
22	80.5	10.8	1036	7	US-11-096-568A-28315
23	80.5	10.8	1070	7	US-11-096-568A-28314
24	80.5	10.8	1276	7	US-11-096-568A-28313
25	79.5	10.6	943	6	US-10-475-204-34

#### ALIGNMENTS

##### RESULT 1

US-10-793-626-652  
; Sequence 652, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 652  
; LENGTH: 746  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-652

Query Match	14.7%	Score 110;	DB 6;	Length 746;
Best Local Similarity	27.0%	Pred. No. 0.09;		
Matches	43;	Conservative	22;	Mismatches 52;
				Indels 42;
				Gaps 8;

  

QY	5	VTIIONGKMSSTIVSEEDFILPVYK	-----GLEKGYQPDG---EISGFE-----G 48
DB	596	ITIGNKQIKQOSKSGTKVLPHSKVLMTDGELTMP	---DMTGWTKEDVLAFEDLTKLKVS 654
QY	49	KKDAGYVNI--LSKDTFLPKVPFKIEERKEBENKPTDVS	-----KKKNPQVNSQLNES 102
DB	655	TKGNFVTVNQISIKQIILK	-----NKKIEVLSAEDTDDQKTDSDSDN 701
QY	103	HRKEDLQREHSQKSDSTKQVTVTLQKNISSTNN 141	
DB	702	KSKKDKADEHSNTSSSTKN	-----DKSNADSKNDSDD 734

##### RESULT 2

US-11-189-817-2  
; Sequence 2, Application US/11189817  
; Publication No. US20060030006A1  
; GENERAL INFORMATION:  
; APPLICANT: INSTITUT PASTEUR  
; APPLICANT: DRUILHE, PIERRE  
; TITLE OF INVENTION: IDENTIFICATION OF A CONSERVED REGION OF PLASMODIUM FALCIPARUM





```

QY 6 TIQNGKMSSTINSEEDFILPVYGELE-KYQF-----DGVHISPEGKK----- 50
Db 355 SVNNESNMDTFVRH-----PIKTGMLNGKKYMYMETTNDDDYWKDFMVEGQVETISKDA 409
QY 51 -----DAGVINL-SKDTFIKVPFKKIEKKKEE 78
Db 410 KNNTRTIIPVVEGKTIYDAIVKHVKTIDYDQYHVRIVDKCAFTKANTDKSNKKQQD 469
QY 79 NKPTFDV-----SKKDNQPNVHSQLNESHRKEDLQ-----REEHSOKSDSTKQVT-ATVL 128
Db 470 NSAKKEATPATPSKPTTSPVEKSKQKQDSQKNDKQLPSVEKENDASSESGDKTPTATKP 529
QY 129 DKNNISSKSTNNPNK 144
Db 530 TKGEVESSTT--FTK 543

RESULT 6
US-11-096-568A-4771
; Sequence 4771, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 4771
; LENGTH: 140
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(140)
; OTHER INFORMATION: Ceres Seq. ID no. 14304111
US-11-096-568A-4771

Query Match 11.6%; Score 87; DB 7; Length 140;
Best Local Similarity 25.9%; Pred. No. 1.1;
Matches 36; Conservative 28; Mismatches 43; Indels 32; Gaps 7;

QY 10 GKMSSTI-----VSEEDFILPVY--KGELEKGYQPDGWEISGEGKKDAGVINLSKDT 62
Db 8 GKQASAREDHGSGEEDKIPAYRRGRPQKPMKD-----FEEDEEE----- 52
QY 63 FIKPVFKKIEKKKEENKPTFDVSKKDNQ-----VHSQLNESHRKEDLQREHSOKSD 118
Db 53 --BELVKEEEEEEDDSVT---SKBENERKRWINGSNTDANEKENGLSKSRDGS 107
QY 119 STKQVATATVLDKNNISSKS 137
Db 108 STKS-TSTGRQNGRRRS 125

RESULT 7
US-11-188-298-1015
; Sequence 1015, Application US/11188298
; Publication No. US20060075522A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 1015
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Thermococcus kodakaraensis

```



; TYPE: PRT  
; ORGANISM: Borrelia burgdorferi  
US-11-196-475-76

```

Query Match      11.4%; Score 85; DB 7; Length 708;
Best Local Similarity 27.6%; Pred. No. 12;
Matches 43; Conservative 22; Mismatches 53; Indels 38; Gaps 8;

Qy 3 VVTVIQNGKEMSSITVSEEDPILPVYKGELEKGYQFDGWEISGPEGKKDAGYVNLK-- 60
   : ||| : : || : : || : : || : : || : : || : : || : : ||
Db 258 ITETIENLRDLEKATDEE-----HKKEIES--QVDAKKKBELEDKA---INLDKAQ 306

Qy 61 -----DTPIKPVFKIIEKKGSEENKP-TFDVSKKDNQPVNHSQNLNESHKK 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 QKLDSAEDNLDVQDVTREKIQEDINEINEKNLKPFGDVSSPKVDQL---QIKES--L 361

Qy 106 EDLQRE-----EHSQKSDSTKDVTVATVLDKKNISKK 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 EDLQEQLEKAGDENQKREIKQIBIKKRDRELLASK 397
   |||| : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
US-10-793-626-2058
; Sequence 2058, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2058
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2058

```

[illegible]

```

RESULT 14
US-11-052-554A-83
; Sequence 83, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052.554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763

```

```

; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 83
; LENGTH: 1694
; TYPE: prt
; ORGANISM: Haemophilus influenzae
US-11-052-554A-83

```

```

Query Match      11.3%; Score 84.5; DB 7; Length 1694;
Best Local Similarity 26.1%; Pred. No. 36;
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

Qy 56 INLSKDTFTKPVFKIEKKKEENKCTFDVSKKQNPQVNHSQLNSHRRKEDLQREHHS- 114
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1288 INTGSAITATETABKSKDKQTETAASTETDASQKANTVADNSVANNSESDDPKSRRRSI 1347
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 115 --QKSDSTKDTATVLDKKNISKSTTNPNK 144
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 1348 SQPQETASBETTAASTDETTIADNSKRSKPNR 1379
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 15
US-11-096-568A-3070
; Sequence 3070, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fr
; TITLE OF INVENTION: 3070
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3070
; LENGTH: 493
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 15172486
US-11-096-568A-3070

```

Query Match	11.2%	Score 83.5;	DB 7;	Length 493;
Best Local Similarity	25.7%;	Pred. No. 10;		
Matches	43;	Conservative 21;	Mismatches 62;	Indels 41; Gaps 8;
QY	6	TIQNGEM--SSTIVSEB-----	DFLLPVYKGELEKGYQPGWEISGEGGKKDAG	53
DB	80	TOSSGKQTADANTIVTEKKKPGKVPKKIKTPVSK---	KQETADSNKTETLSDKKDEG	136
QY	54	YVI-----	NLSKDTFIKVPKKTIEE--KCBENKPTFDVSKKONPQVNH	97
DB	137	NVVAVOAQDDTQSTGQTANADTTVTPEVKTKGKVPKQSKTPT---	SEKRDN--TADS	191
QY	98	QLNESHKEDLQREHSQKSDSTKOVTAATVLDRKNISGKSTTNNPK		144
DB	192	SKTETKSDKDDKKEER-----VTGEKSGAKTKLKASDKDVTNVGK		233

```

RESULT 16
US-11-096-568A-3071
; Sequence 3071, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-determined DN
; TITLE OF INVENTION: Tiberby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3071
; LENGTH: 493

```

```
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(493)
; OTHER INFORMATION: Ceres Seq. ID no. 16625552
US-11-096-568A-3071

Query Match
Best Local Similarity 11.2%; Score 83.5; DB 7; Length 493;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

Qy 6 TIONKEM--SSTIVSE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 53
Db 80 TQSSGQTADANTIVTEKKPKGVVKKIKTPVSK---KKDETADSNKTETLSKKDEG 136
Qy 54 YVI-----NLSKDTFKVPFKIEE---KKEENKPTFDVSKKKNPQVNH 97
Db 137 NVVAVQAQDDTQSTGQTANADTTVPEVKTKGVVKKQSKTPT---SEKRDN--TADS 191
Qy 98 QLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 144
Db 192 SKTETSKDKDKKEER-----VTGEKSGAKTDKLNKASDKDVTNVKKG 233

RESULT 17
US-11-096-568A-3069
; Sequence 3069, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592FUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 3069
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc_feature
; LOCATION: (1)..(510)
; OTHER INFORMATION: Ceres Seq. ID no. 15172485
US-11-096-568A-3069

Query Match
Best Local Similarity 11.2%; Score 83.5; DB 7; Length 510;
Matches 43; Conservative 21; Mismatches 62; Indels 41; Gaps 8;

Qy 6 TIONKEM--SSTIVSE-----DFILPVYKGELEKGYQPDGWEISGFEGKKDAG 53
Db 97 TQSSGQTADANTIVTEKKPKGVVKKIKTPVSK---KKDETADSNKTETLSKKDEG 153
Qy 54 YVI-----NLSKDTFKVPFKIEE---KKEENKPTFDVSKKKNPQVNH 97
Db 154 NVVAVQAQDDTQSTGQTANADTTVPEVKTKGVVKKQSKTPT---SEKRDN--TADS 208
Qy 98 QLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 144
Db 209 SKTETSKDKDKKEER-----VTGEKSGAKTDKLNKASDKDVTNVKKG 250

RESULT 18
US-11-188-298-15964
; Sequence 15964, Application US/11188298
; Publication No. US2006007552A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
```

```
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,976
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 15964
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus DSM 3638
US-11-188-298-15964

Query Match
Best Local Similarity 11.0%; Score 82; DB 7; Length 439;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

Qy 35 GYQFDGWEISGFEGKKDAGVYVNLKDTPT-----KP-----VFK 69
Db 45 GISPDGSSVFGQIEDSLVFKADPTTYVEVPMDNVARVGYFYKONKPYGADPRGLK 104
Qy 70 KIEKKEEEN-----KPTFDVSKKKN----- 91
Db 105 RALELEKEGKYAVIGPEPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIAYM 164
Qy 92 -----POVNHSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 133
Db 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKADNIVSFKYITKAVAEHGL 215

RESULT 19
US-11-188-298-16606
; Sequence 16606, Application US/11188298
; Publication No. US2006007552A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: GENES AND USES FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53452)B
; CURRENT APPLICATION NUMBER: US/11/188,298
; CURRENT FILING DATE: 2005-07-22
; PRIOR APPLICATION NUMBER: 60/592,978
; PRIOR FILING DATE: 2004-07-31
; NUMBER OF SEQ ID NOS: 22569
; SEQ ID NO 16606
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Pyrococcus woesei
US-11-188-298-16606

Query Match
Best Local Similarity 11.0%; Score 82; DB 7; Length 439;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

Qy 35 GYQFDGWEISGFEGKKDAGVYVNLKDTPT-----KP-----VFK 69
Db 45 GISPDGSSVFGQIEDSLVFKADPTTYVEVPMDNVARVGYFYKONKPYGADPRGLK 104
Qy 70 KIEKKEEEN-----KPTFDVSKKKN----- 91
Db 105 RALELEKEGKYAVIGPEPEFYLFKNGTWELEIPDVGGYFDILTLDKARDIRREIAYM 164
Qy 92 -----POVNHSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 133
Db 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKADNIVSFKYITKAVAEHGL 215

RESULT 20
US-11-098-686-10232
; Sequence 10232, Application US/11098685
; Publication No. US20060024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; TITLE OF INVENTION: FROM LAWSONIA INTRACELLULARIS AND METHODS OF USING
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
```

; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: PCT/US03/31318  
; PRIOR FILING DATE: 2003-10-01  
; PRIOR APPLICATION NUMBER: US 60/416,395  
; PRIOR FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 11433  
; SOFTWARE: PatSeq for Windows Version 4.0  
; SEQ ID NO 10232  
; LENGTH: 8746  
; TYPE: PRT  
; ORGANISM: Lawsonia intracellularis  
US-11-098-686-10232

Query Match 10.9%; Score 81.5; DB 7; Length 8746;  
Best Local Similarity 25.9%; Pred. No. 4.6e+02;  
Matches 38; Conservative 24; Mismatches 60; Indels 25; Gaps 7;  
QY 12 EMSSTIVSEEDFIL--PVYKGLKGYQFDGWE--ISGFEGKDGAGYVINLSKDTFFIKPV 67  
Db 3358 EGEDVAEEEDLILLWDILDNEAEGTKEHAEVKGVEGVFDG-----ISED--KP- 3410  
QY 68 FKIKBEKKEENKPTFDVSKKONPQVNHSQLNESHKEDL-----QREHSQKS 117  
Db 3411 --KDDKEQEOKATLGDSSGRTIEESQPPQEEBEKKENSFGSNGESPSPOEESYDE 3468  
QY 118 DSTKDVATVLDKNNISSKSTNNPNK 144  
Db 3469 TSSVVTSPLLSINEV--KOTEDKSAK 3493

## RESULT 21

US-10-793-626-264  
; Sequence 264, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P034800S  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patent Ver. 2.1  
; SEQ ID NO 264  
; LENGTH: 785  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURES:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-264

Query Match 10.8%; Score 81; DB 6; Length 785;  
Best Local Similarity 24.7%; Pred. No. 29;  
Matches 37; Conservative 27; Mismatches 66; Indels 20; Gaps 7;  
QY 4 TWTIONGKEMS-STIVSEEDFILPVYKGLKGYQFDGWEISGFEGKDGAGY-VINLSKD 61  
Db 638 TITEENGIVICITTVTGEEDLMVVTNAGVI---IFLDVHDISQ-NGRAAQGVRLMKLGDG 693  
QY 62 TFIKPVFKKIEKKEEN-----KPTFDVSKKONPQVNHSQLNESHKEDLQREH 112  
Db 694 QFVSTVAKVNEEDDNEENADRAQSTTTTETADVEEVVD----DQTPGNAIHTEGDAMES 749  
QY 113 -HSQKSDSTKVTATVLDKNNISSKSTNN 141  
Db 750 VEPFENDRIDIRQDFMDRVNEDIESADN 779

## RESULT 22

US-11-096-568A-28315  
; Sequence 28315, Application US/11096568A

; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nickolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28315  
; LENGTH: 1036  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURES:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1036)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712010  
US-11-096-568A-28315

Query Match 10.8%; Score 80.5; DB 7; Length 1036;  
Best Local Similarity 24.6%; Pred. No. 44;  
Matches 34; Conservative 27; Mismatches 64; Indels 13; Gaps 5;  
QY 9 NGKEMSTIVSEEDFILPVYKGL-EGYQFDGWEISGFEGKDGAGYVINLSKDTF---- 63  
Db 516 NEQTPAKTIIVKKK--IIRKVAKRKVAEIDNKMVG--DSKKDGDSDKKVMEVGVKSSDGS 572  
QY 64 --IKPVFKKIEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREH---SOKSD 118  
Db 573 VEMKPTAESLEDVKDENASKTVDVKQETGSPDTKKKGASSSSKKDTGTGEDKKAERKNN 632  
QY 119 STKDVATVLDKNNISSK 136  
Db 633 SETMSEKKIDRNNNTDEK 650

## RESULT 23

US-11-096-568A-28314  
; Sequence 28314, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nickolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 28314  
; LENGTH: 1070  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURES:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(1070)  
; OTHER INFORMATION: Ceres Seq. ID no. 2712009  
US-11-096-568A-28314

Query Match 10.8%; Score 80.5; DB 7; Length 1070;  
Best Local Similarity 24.6%; Pred. No. 46;  
Matches 34; Conservative 27; Mismatches 64; Indels 13; Gaps 5;  
QY 9 NGKEMSTIVSEEDFILPVYKGL-EGYQFDGWEISGFEGKDGAGYVINLSKDTF---- 63  
Db 550 NEQTPAKTIIVKKK--IIRKVAKRKVAEIDNKMVG--DSKKDGDSDKKVMEVGVKSSDGS 606  
QY 64 --IKPVFKKIEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREH---SOKSD 118  
Db 607 VEMKPTAESLEDVKDENASKTVDVKQETGSPDTKKKGASSSSKKDTGTGEDKKAERKNN 666  
QY 119 STKDVATVLDKNNISSK 136  
Db 667 SETMSEKKIDRNNNTDEK 684

```
RESULT 24
US-11-096-568A-28313
; Sequence 28313, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 28313
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc feature
; LOCATION: (1)..(1276)
; OTHER INFORMATION: Ceres Seq. ID no. 2712008
US-11-096-568A-28313

Query Match          10.8%; Score 80.5; DB 7; Length 1276;
Best Local Similarity 24.6%; Pred. No. 56;
Matches 34; Conservative 27; Mismatches 64; Indels 13; Gaps 5;

QY 9 NGKMSSTIVSEDFILPVYKGL-EGYQFDGWEISGFEKGKQAGYVINLSKDTF---- 63
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 756 NEQTPAKTIVKK-IIRVAKKVAEIDNWDG--DSKKDGSDEKKYMEVGKSSDSGS 812
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 64 --IKPVFKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREH---SQKSD 118
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 813 VMKPTAESLDVDENASKTVDVKQETGSPDTKKKGASSSSKDTTKTGEDKKAQKN 872
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 119 STPKDVTATVLDKNNISSK 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 873 SETMSEGGKIDRNWTEK 890
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 25
US-10-475-204-34
; Sequence 34, Application US/10475204
; Publication No. US20050277116A1
; GENERAL INFORMATION:
; APPLICANT: PRESIDENT AND FELLOWS OF HARVARD COLLEGE
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE IDENTIFICATION OF
; FILE REFERENCE: HMV-056.25
; CURRENT APPLICATION NUMBER: US/10/475,204
; CURRENT FILING DATE: 2003-10-17
; PRIOR APPLICATION NUMBER: PCT/US02/13008
; PRIOR FILING DATE: 2002-09-27
; PRIOR APPLICATION NUMBER: 60/285,509
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 34
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-475-204-34

Query Match          10.6%; Score 79.5; DB 6; Length 943;
Best Local Similarity 23.1%; Pred. No. 48;
Matches 36; Conservative 16; Mismatches 55; Indels 49; Gaps 5;

QY 26 PVYKGELEKGYQFDGWEISGFEKGKQAGYVINLSKDTTKPVFKIEBK-----EEN 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 375 PSDKTVLDTSYALIDETVNNYRSTKYEMYSKNAEKPSRSKRTIKQQRKFMKAPAEQ- 433
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 80 KPTFDVSKKCD-NPQVNHSQLNESHKEDLQREH-----SQKSDSTK 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 434 ---LDVQSQDENIHTSHITQDEFCRNSDRNWEHEENGNDVCYKKQMPVGVGSKSSTRK 490
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 D-----VTATVLDKNNISSK 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 491 DKESKKKRFSSSESKNKLVPPEVTTSTVTGRRISRR 526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 26
US-10-793-626-2482
; Sequence 2482, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2482
; LENGTH: 568
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2482

Query Match          10.6%; Score 79; DB 6; Length 568;
Best Local Similarity 24.8%; Pred. No. 29;
Matches 33; Conservative 22; Mismatches 44; Indels 34; Gaps 6;

QY 9 NGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEKGKQAGYVINLSKDTFKIPVF 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 NXHAVNEKILNQEDYDVVPSLISEL---YQYQ-TSISN-----TYVL-----TFSDEVI 273
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 69 KKTBEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQLKSDTKDVTATVL 128
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 KALNKKIENESTPLFEAA-----\NHVQVNELSDENEDRYEYDR-----YI 315
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 129 DKNNISSKSTNN 141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 ELNTLKDSLTSN 328
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 27
US-11-196-475-68
; Sequence 68, Application US/11196475
; Publication No. US20050271682A1
; GENERAL INFORMATION:
; APPLICANT: Dattwyler, Raymond J.
; APPLICANT: Gomes Solecki, Maria J. C.
; APPLICANT: Luft, Benjamin J.
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: Recombinant Constructs of Borrelia
; FILE REFERENCE: 2631.1001-011
; CURRENT APPLICATION NUMBER: US/11/196,475
; CURRENT FILING DATE: 2005-08-03
; PRIOR APPLICATION NUMBER: US 08/148,191
; PRIOR FILING DATE: 1993-11-01
; PRIOR APPLICATION NUMBER: US 08/235,836
; PRIOR FILING DATE: 1994-04-29
; PRIOR APPLICATION NUMBER: US 09/666,017
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/226,484
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: PCT/US01/24736
; PRIOR FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 213
```







Db 170 NIQAQSKETEEAEELKMLKRLSLNFKATPMPSEFYKEPPPPKVELKIKPTTRARSPKLGSRKN 229  
QY 133 ISSKSTTNNPN 143  
Db 230 TSSGGTEGNPN 240  
RESULT 35  
US-11-096-568A-17851  
; Sequence 17851, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 17851  
; LENGTH: 386  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(386)  
; OTHER INFORMATION: Ceres Seq. ID no. 12361231  
US-11-096-568A-17851  
Query Match 10.5%; Score 78.5; DB 7; Length 386;  
Best Local Similarity 19.4%; Pred. No. 20;  
Matches 37; Conservative 22; Mismatches 67; Indels 65; Gaps 4;  
QY 3 VTVTIQNGKMSSTIVSRDPLPVYKGELEKGYQDGEWISGFGKGDAGVVLNSKDT 62  
Db 124 VDCSLNGDAGSKKKAERKSPRPVAK-----ETPSLDSNEKKKTQKASQNH 171  
QY 63 FIKPVPKIEEKKEBENKP-----TPDVSKKKNPNVHISQLNESHKEDLQRE 111  
Db 172 SVK---KQIEESNESVQPVGVSTPSYGFSPKCDERAEKREFYSKLEEKIHAQLEKS 228  
QY 112 EHSQSDSTKVQVATVLDKN-----N 132  
Db 229 NIQAQSKETEEAEELKMLKRLSLNFKATPMPSEFYKEPPPPKVELKIKPTTRARSPKLGSRKN 288  
QY 133 ISSKSTTNNPN 143  
Db 289 TSSGGTEGNPN 299  
RESULT 36  
US-10-793-626-1780  
; Sequence 1780, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1780  
; LENGTH: 1155  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-1780

Query Match 10.5%; Score 78.5; DB 6; Length 1155;  
Best Local Similarity 27.3%; Pred. No. 74;  
Matches 24; Conservative 12; Mismatches 27; Indels 25; Gaps 2;  
QY 73 EKKEENKPTFDVSKKKN-----PQVNHSQLNESHKEDLQREHSQK 116  
Db 100 EVKVEAPTTSVSKPKANEAVVTNESTPKTKTTRAPTNEESIARTPKTSTTQODSTSKN 159  
QY 117 SDSTKVQVATVLDKNKNISSKSTTNNPNK 144  
Db 160 NPSLKD-----NLNSSSTTSKSK 178  
RESULT 37  
US-10-860-601-5  
; Sequence 5, Application US/10860601  
; Publication No. US20050273871A1  
; GENERAL INFORMATION:  
; APPLICANT: Asano, Yoshihiro  
; APPLICANT: Takashima, Seiji  
; APPLICANT: Kitakaze, Masafumi  
; TITLE OF INVENTION: Method for Diagnosing Arrhythmogenic Right Ventricular Dysplasia  
; FILE REFERENCE: 2144.0210000  
; CURRENT APPLICATION NUMBER: US/10/860,601  
; CURRENT FILING DATE: 2004-06-04  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 5  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: RVAP27  
US-10-860-601-5  
Query Match 10.4%; Score 78; DB 6; Length 191;  
Best Local Similarity 23.9%; Pred. No. 9.6;  
Matches 27; Conservative 29; Mismatches 37; Indels 20; Gaps 4;  
QY 14 SSTIVSRDFTLP-----VYKGELEKGYQPDGW--HISGEGKGDAGVVLNSKDTIK 65  
Db 11 SSSSEDEEYVYKVLDRRVYKGVQVYLLKWKGFSEHNTWPEK-----NLDCPELIS 64  
QY 66 PVFKIKIEKEENKPTFDVSKKKNPNVHISQLNESHKEDLQREHSQKSD 118  
Db 65 EFMKKYKMKKEGEN-----NKPREKSESNKRNKSNFNSADDIKSKKKEQSN 111  
RESULT 38  
US-10-485-517-239  
; Sequence 239, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, Simon  
; APPLICANT: Mond, James  
; TITLE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629WO  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; CURRENT FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118825.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 239  
; LENGTH: 405  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-239

```

Query Match      10.4%; Score 77.5; DB 6; Length 405;
Best Local Similarity 29.7%; Pred. No. 26;
Matches 22; Conservative 16; Mismatches 33; Indels 3; Gaps 2;

69 KKIEEKKEEENKPTFDVSKKONPOVNHSQLN-BSHRKEDIQRSEHSOKSPDKDVTATV 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
51 KAAESTNKELENAITTSASDNQSSDKVDMMQLNQEDNTKNDKQKEMVSSQGNETTSNGKL 110
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

128 LDKNNISSKSTNN 141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
111 IEKGSV--QSTTG 122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:33:35 ; Search time 72.1611 Seconds  
(without alignments)  
833.793 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773  
Perfect score: 748  
Sequence: 1 HRVTVTIQNGKEMSTIVSE.....ATVLDKNNISSKSTNNPNK 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	748	100.0	773	4	US-10-067-385-8
2	748	100.0	2119	3	US-09-769-744A-28
3	748	100.0	2140	4	US-10-282-122A-73670
4	748	100.0	2140	5	US-10-472-928-1180
5	745	99.6	637	3	US-10-617-320-3169
6	615	82.2	117	3	US-09-765-272-68
7	615	82.2	117	6	US-11-106-649-68
8	110	14.7	778	4	US-10-724-972A-5663
9	108	14.4	188	5	US-10-691-672A-7
10	106.5	14.2	470	5	US-10-739-930-6262
11	105	14.0	647	5	US-10-691-672A-3
12	104	13.9	665	3	US-09-820-843A-107
13	103.5	13.8	169	5	US-10-691-672A-2
14	102.5	13.7	775	4	US-10-282-122A-70721
15	101.5	13.6	564	6	US-11-097-143-12723
16	99.5	13.3	707	4	US-10-282-122A-52942
17	95	12.7	973	5	US-10-732-923-18783
18	93.5	12.5	898	4	US-10-425-115-205148
19	93	12.4	869	4	US-10-437-963-122282
20	93	12.4	948	5	US-10-732-923-4286
21	93	12.4	1529	5	US-10-732-923-8762
22	93	12.4	3127	5	US-10-732-923-22588
23	92.5	12.4	540	5	US-10-732-923-22820
24	92.5	12.4	2468	4	US-10-755-889-615
25	92.5	12.4	2468	5	US-10-489-740-216
26	92.5	12.4	2519	5	US-10-450-763-46995
27	92	12.3	1005	4	US-10-437-963-187665

28	91.5	12.2	903	4	US-10-282-122A-52328	Sequence 52328, A
29	91	12.2	2060	4	US-10-381-596A-2	Sequence 2, Appli
30	90	12.0	932	4	US-10-282-122A-52510	Sequence 52510, A
31	89.5	12.0	511	4	US-10-289-762-509	Sequence 509, App
32	89.5	12.0	1408	6	US-11-097-143-2904	Sequence 2904, App
33	89	11.9	519	4	US-10-437-963-160737	Sequence 160737, A
34	89	11.9	645	4	US-10-282-122A-70294	Sequence 70294, A
35	89	11.9	645	5	US-10-470-048B-414	Sequence 414, App
36	89	11.9	654	4	US-10-172-502-10	Sequence 10, Appl
37	89	11.9	654	6	US-11-020-509-10	Sequence 10, Appl
38	88.5	11.8	281	6	US-11-097-143-21276	Sequence 21276, A
39	88	11.8	225	4	US-10-032-585-7829	Sequence 7829, Ap
40	88	11.8	645	5	US-10-470-048B-142	Sequence 142, App
41	88	11.8	815	5	US-10-496-905-23	Sequence 23, Appl
42	88	11.8	1980	5	US-10-482-834A-144	Sequence 144, App
43	87.5	11.7	441	5	US-10-510-812-14	Sequence 14, Appl
44	87.5	11.7	717	5	US-10-732-923-23071	Sequence 23071, A
45	87.5	11.7	1196	4	US-10-282-122A-52737	Sequence 52737, A

## ALIGNMENTS

## RESULT 1

US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score 748;	DB 4;	Length 773;
Best Local Similarity	100.0%	Pred. No. 7.8e-58;		
Matches 144;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	HRVTVTIQNGKEMSTIVSEDFILPVYKGLRGYQPDGWEISGFEGKDGAGYVNL	SK 60	
Db	630	HRVTVTIQNGKEMSTIVSEDFILPVYKGLRGYQPDGWEISGFEGKDGAGYVNL	SK 689	
Qy	61	DTFIPKVPKLEEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSD	120	
Db	690	DTFIPKVPKLEEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSD	749	
Qy	121	KDVTATVLDKNNISSKSTNNPNK	144	
Db	750	KDVTATVLDKNNISSKSTNNPNK	773	

## RESULT 2

US-09-769-744A-28  
; Sequence 28, Application US/09769744A  
; Publication No. US20030134407A1  
; GENERAL INFORMATION:  
; APPLICANT: Le Page, Richard WF  
; APPLICANT: Wells, Jeremy M  
; APPLICANT: Hanniffy, Sean B  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins

```
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match          100.0%; Score 748; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 2.7e-57;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 60
|
Db 1942 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 2001
|
Qy 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKREDLQREHSQKSDST 120
|
Db 2002 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKREDLQREHSQKSDST 2061
|
Qy 121 KDVTATVLDKNNISSKSTTNNPNK 144
|
Db 2062 KDVTATVLDKNNISSKSTTNNPNK 2085
|
```

## RESULT 3

```
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match          100.0%; Score 748; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.7e-57;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 60
|
Db 1963 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 2022
|
Qy 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKREDLQREHSQKSDST 120
|
Db 2023 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKREDLQREHSQKSDST 2082
|
Qy 121 KDVTATVLDKNNISSKSTTNNPNK 144
|
Db 2083 KDVTATVLDKNNISSKSTTNNPNK 2106
|

RESULT 4
US-10-472-928-1180
; Sequence 1180, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (O.E+01)
US-10-472-928-1180

Query Match          100.0%; Score 748; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 2.7e-57;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 60
|
Db 1963 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 2022
|
Qy 61 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKREDLQREHSQKSDST 120
|
Db 2023 DTFIKPVFKKIEKKKEENKPTFDVSKKONPVNHSQLNESHKREDLQREHSQKSDST 2082
|
Qy 121 KDVTATVLDKNNISSKSTTNNPNK 144
|
Db 2083 KDVTATVLDKNNISSKSTTNNPNK 2106
|

RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
```

Publication No. US20050136404A1  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
NUMBER OF SEQUENCES: 5206  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: <Unknown>  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: <Unknown>  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/617,320  
FILING DATE: 10-Jul-2003  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,433  
FILING DATE: 30-Jun-1998  
APPLICATION NUMBER: 60/085131  
FILING DATE: May 12, 1998  
APPLICATION NUMBER: 60/051553  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Denke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3169:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 637 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...637  
SEQUENCE DESCRIPTION: SEQ ID NO: 3169:  
US-10-617-320-3169  
Query Match 99.6%; Score 745; DB 5; Length 637;  
Best Local Similarity 99.3%; Pred. No. 1.1e-57;  
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HRVTVIQNGKEMSTIIVSEEDFILPVYKGBLEKGYQFDGWEISGFEKGDAGYVINLSK 60  
DB 460 HRVTVIQNGKEMSTIIVSEEDFILPVYKGBLEKGYQFDGWEISGFEKGDAGYVINLSK 519  
QY 61 DTFIKPVFKIEEKKKEENKPTFDVSKKDPQVNHSQLNESHKEDLQREHSQKSDST 120  
DB 520 DTFIKPVFKIEEKKKEENKPTFDVSKKDPQVNHSQLNESHKEDLQREHSQKSDST 579  
QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
DB 580 KDVTATVLDKNNISSKSTNNPNK 603  
RESULT 6  
US-09-765-272-68  
Sequence 68, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:

APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-765-272-68  
Query Match 82.2%; Score 615; DB 3; Length 117;  
Best Local Similarity 100.0%; Pred. No. 5.3e-47;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 28 YKGBLEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIEEKKKEENKPTFDVSK 87  
DB 1 YKGBLEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFKIEEKKKEENKPTFDVSK 60  
QY 88 KQNPQVNHSQLNESHKEDLQREHSQKSDSTQVATVLDKNNISSKSTNNPNK 144  
DB 61 KQNPQVNHSQLNESHKEDLQREHSQKSDSTQVATVLDKNNISSKSTNNPNK 117  
RESULT 7  
US-11-106-649-68  
Sequence 68, Application US/11106649  
Publication No. US20050181439A1  
GENERAL INFORMATION:  
APPLICANT: Choi et al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
FILE REFERENCE: PB340P2C3D1  
CURRENT APPLICATION NUMBER: US/11/106,649  
CURRENT FILING DATE: 2005-04-15  
PRIORITY APPLICATION NUMBER: US 09/765,271  
PRIORITY FILING DATE: 2001-01-22  
PRIORITY APPLICATION NUMBER: US 09/536,784  
PRIORITY FILING DATE: 2000-03-28  
PRIORITY APPLICATION NUMBER: US 08/961,083  
PRIORITY FILING DATE: 1997-10-30  
PRIORITY APPLICATION NUMBER: US 60/029,960  
PRIORITY FILING DATE: 1996-10-31  
NUMBER OF SEQ ID NOS: 454  
SOFTWARE: Patent in version 3.3  
SEQ ID NO 68

```
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match      82.2%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.3e-47;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 YKGELEKGYQPGWISFEGKDKAGYVNLKDTFKIPVFKIIEKKKEENKPTFDVSK 87
Db 1 YKGELEKGYQPGWISFEGKDKAGYVNLKDTFKIPVFKIIEKKKEENKPTFDVSK 60

QY 88 KKDNPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPNK 144
Db 61 KKDNPQVNHSQLNESHKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPNK 117

RESULT 8
US-10-724-972A-5663
; Sequence 5663, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 5663
; TYPE: PRT
; ORGANISM: S.epidermidis
US-10-724-972A-5663

Query Match      14.7%; Score 110; DB 4; Length 778;
Best Local Similarity 27.0%; Pred. No. 0.44;
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;

QY 5 VTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQFGW---EISGFE-----G 48
Db 628 ITIINGKQIKQOSVSKGTVLPHSKVLMWTDGELTMP-DMTGWTYKEDVLAFEDLTIKIVS 686

QY 49 KKDAGYVIN--LSKDTFKIPVFKIIEKKKEENKPTFDVS-----KKDNPQVNHSQLNES 102
Db 687 TKGNGFVNTQSIKQQLIK-----NKKDIEVLSAEDTDDQKTDSDSDN 733

QY 103 HKKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPNK 141
Db 734 KSKDKDAEDHSHNTSSSTKN-----DKSNADSKNDSDD 766

RESULT 9
US-10-691-672A-7
; Sequence 7, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
```

```
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 7
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; NAME/KEY: SITE
; LOCATION: (1)..(188)
; OTHER INFORMATION: MSP3a to MSP3f
US-10-691-672A-7

Query Match      14.4%; Score 108; DB 5; Length 188;
Best Local Similarity 23.2%; Pred. No. 0.12;
Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 7 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFGWISGFI--EGKKDAG-----YVINLS 59
Db 15 VLKAKEASS-----YDYIL-----GWFFGGGVPEHKKEENMLSHLYVSSKD 55

QY 60 KDTFKIPVFKIIEKKKE-----ENKPTFDVSKKKNPQVNHSQLNESHKKE 106
Db 56 KENISKENDVDLDEKKEAEETEEELKKNKEETEEISEDEEEEEEKKEEEDKKK 115

QY 107 DLQREHSHQKSDSTKDVATVLDKNNISSKSTNN 141
Db 116 EOEKEQSNENDQKQMEAE-----QNLISKNNNN 145

RESULT 10
US-10-739-930-6262
; Sequence 6262, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 6262
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C271270_1.p
US-10-739-930-6262

Query Match      14.2%; Score 106.5; DB 5; Length 470;
Best Local Similarity 20.3%; Pred. No. 0.49;
Matches 36; Conservative 31; Mismatches 65; Indels 45; Gaps 5;

QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFGWISGFE-----GKDDA 52
Db 84 NRVTDTVQNNNGSK-----YVQDLARRIRYDE-EATGSQSAQRIDHPNQNV 131

QY 53 GYVINLSKDTFKIPVFKIIEKKKEENKPTFDVSKKKN-----PQ 93
Db 132 GITEKAFENSPIEETSHRVDDNKINQKNFTAAKSSSENASVRSVFGADHKRAEYMGKPM 191

QY 94 VNHSQANE-----SHRKEDLQREHSHQKSDSTKDVATVLDKNNISSKSTNNPNK 144
Db 192 ENRDQVRQTSAEKSHRKNVTKSEKPRDQGVKKTEAKDKRKNKEKKEETESINK 248

RESULT 11
US-10-691-672A-3
; Sequence 3, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
```

```
; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 3
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
; NAME/KEY: SITE
; LOCATION: (1)..(647)
; OTHER INFORMATION: GLURP MSP3 fusion protein
US-10-691-672A-3

Query Match      14.0%; Score 105; DB 5; Length 647;
Best Local Similarity 22.7%; Pred. No. 0.98;
Matches 42; Conservative 35; Mismatches 52; Indels 56; Gaps 9;

QY 1 HRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELEK-----GYQ 37
DB HEETVSQSNPEKADNGVNSNNELNEFV-----ESEKSEHARSKAKEASSYD 485
QY 38 FD-GWEISGF--EGKKDAG-----YVINLSKDTPIKPVFKIEKKKEENKPTFDVSKK 89
DB YILGWEFGGVPKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEAEETEEHEELE 544
QY 90 DNPQVNHSQLN-----ESHKRDQLQREHSQKSDSKDTKVATVLDKNNISSK 136
DB KNEERETSEISEDEREEERKEEENKKEQKEQESQNNENDQKKDMEA-----QNLISK 599
QY 137 STTN 141
DB 600 NQNN 604

RESULT 12
US-09-820-843A-107
; Sequence 107, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|3845248
US-09-820-843A-107

Query Match      13.9%; Score 104; DB 3; Length 665;
Best Local Similarity 25.9%; Pred. No. 1.2;
Matches 42; Conservative 29; Mismatches 35; Indels 56; Gaps 10;

QY 1 HRVT-VTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGFEGKDGAGYVIN 57
DB HRQNELNLQSGK-----NEQDI-----KNKEGKQ-----DLSNSNAENKKD----- 174
QY 58 LSKDTPIKPVFKIEKKE-----EENKPTFD-----VSKKDNQPVNHSQNLNESH 103

; TITLE OF INVENTION: MALARIAL VACCINES CONTAINING IT
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 3
; LENGTH: 647
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
; NAME/KEY: SITE
; LOCATION: (1)..(647)
; OTHER INFORMATION: GLURP MSP3 fusion protein
US-10-691-672A-3

Query Match      14.0%; Score 105; DB 5; Length 647;
Best Local Similarity 22.7%; Pred. No. 0.98;
Matches 42; Conservative 35; Mismatches 52; Indels 56; Gaps 9;

QY 1 HRVTVTIQNGKEM-----SSTIVSEEDFILPVYKGELEK-----GYQ 37
DB HEETVSQSNPEKADNGVNSNNELNEFV-----ESEKSEHARSKAKEASSYD 485
QY 38 FD-GWEISGF--EGKKDAG-----YVINLSKDTPIKPVFKIEKKKEENKPTFDVSKK 89
DB YILGWEFGGVPKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEAEETEEHEELE 544
QY 90 DNPQVNHSQLN-----ESHKRDQLQREHSQKSDSKDTKVATVLDKNNISSK 136
DB KNEERETSEISEDEREEERKEEENKKEQKEQESQNNENDQKKDMEA-----QNLISK 599
QY 137 STTN 141
DB 600 NQNN 604

RESULT 13
US-10-691-672A-2
; Sequence 2, Application US/10691672A
; Publication No. US20050112133A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; TITLE OF INVENTION: GLURP-MSP3 FUSION PROTEIN, IMMUNOGENIC COMPOSITIONS AND
; FILE REFERENCE: 02356.0085
; CURRENT APPLICATION NUMBER: US/10/691,672A
; CURRENT FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 2
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1)..(169)
; OTHER INFORMATION: MSP3 amino acids 212-380
US-10-691-672A-2

Query Match      13.8%; Score 103.5; DB 5; Length 169;
Best Local Similarity 25.2%; Pred. No. 0.26;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 11 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 63
DB 1 KEASS-----YDYIL-----GWFGGVPKKEENMLSHLYVSSKDKENI 41
QY 64 IKPVFKIEKKKEENKPTFDVSKKDNQPVNHSQNLN-----ESHKRDQLQ 110
DB 42 SKENDVDLDE-KEEAEETEEHEELEKKEEETSEISEDEREEERKEEENKKEQKEQ 100
QY 111 EHSQKSDSKDTKVATVLDKNNISSKSTTN 141
DB 101 EQSNENDQKKDMEA-----QNLISKQNQNN 126

RESULT 14
US-10-282-122A-70721
; Sequence 70721, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70721
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-70721

Query Match      13.7%; Score 102.5; DB 4; Length 775;
Best Local Similarity 24.7%; Pred. No. 2; Mismatches 61; Indels 29; Gaps 7;
Matches 38; Conservative 26;

QY 5 VTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW-----EISGFE-----G 48
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 625 ITIIGNGKIQKQSVKSGTKVLPKSHKVMLTGDELTPM-DMTGWTKEVDLAFEDLTPIKIVS 683

QY 49 KKDAGYVYN--LSKDTFFIKPVK-----KIEKKKEENKPTPDVSKKKONPOVNHISQ 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 TKNGGFVTVNQTSKGGIINKDKIEVLSLSAEPTDDQKTDDESDSKSKKDKVDDEDSN 743

QY 99 LNESHRKEDLQREHSQKSDSKDVTATVLDKNN 132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 ASSSKNEKSNADSKNDSDDSTNETSGS--ERNN 775

RESULT 15
US-11-097-143-12723
; Sequence 12723, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12723
; LENGTH: 564
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-12723

Query Match      13.6%; Score 101.5; DB 6; Length 564;
Best Local Similarity 24.5%; Pred. No. 1.7; Mismatches 51; Indels 25; Gaps 5;
Matches 34; Conservative 29;

QY 21 EDFILPVYKGELEKGYQFDGW-----EISGFEKKDAGYVI-----NLSKDTFFIK 65
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 EDLDTPLSESFSK--VFDGWVDEHRDHDGHDVQEPGSEALDDHDEHDDHDDHDEDEE 135

QY 66 PVFKKIEKKKEENKPT-----PDVSKKKONPOVNHISQKSHRKEDLQREHSQKSDS 119
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 PLTELEEELEEEETPEDEPADEFEDEENNA--GENITAEDEEEDEENDD 193

QY 120 TKDVTATVLDKNNISKSST 138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 EGTVEATVEATTEAT 212

RESULT 16
US-10-282-122A-52942
; Sequence 52942, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52942
; LENGTH: 707
; TYPE: PRT
; ORGANISM: Clostridium difficile
; FEATURE:
; NAME/KEY: MISC FEATURE
```



Query Match	13.3%;	Score	99.5;	DB	4;	Length	707;
Best Local Similarity	25.3%;	pred.	No. 3.4;				
Matches	37;	Conservative	26;	Mismatches	60;	Indels	23; Gaps
							4;
Qy	18	VSEEDPLPVYKGELEKGYPDQWEISGFPEGKKDAGVYNLSKTFPKPVPVKKEEK---	74				
	:	:::::	:	:	:	:	:
Db	484	ISIEDDAEGYKEIDSNNQDIGDW---EDKOTTDREYDSNKEDIIEPENKSKCKAKL	540				
	:	:	:	:	:	:	:
Qy	75	-----KEEEN-----KPTFYDKKKONPQVNHSQLNSHREDLQRESHHSOKSD	118				
	:	:	:	:	:	:	:
Db	541	PGFIKONEVEQEENEINDISPDIILDKPVENNQVKSBEIQEONELKE-IKQEPSPQHIE	599				
	:	:	:	:	:	:	:
Qy	119	STKDVTATVLDDKNNISKSTTNPNK	144				

```

RESULT 18
US-10-425-115-205148
; Sequence 205148, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalick, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecu
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,1
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 205148
; LENGTH: 898
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_1
US-10-425-115-205148

```

484	ISIEDDABGVKEELUSDNNQJIGUV	----	EDKUTIDAEYUSNKGDIILEPENKKS	KKKAK	544
QY	75	-----KEEN-----	KPTFDVSKKCNQVNHSQLNSHRRKEDLQR	EESSKSD	118
Db	541	PGFKCNBEVEQEELNDISPDIIIDKPVENNQV	KEEIQELNKE	-IKQEPSPQHIR	599
QY	119	STKDVTATVLDKNNISSKSTTNNPNK	144		

```
Query Match      12.5%; Score 93.5; DB 4; Length 898;
Best Local Similarity 23.5%; Pred. No.15;
Matches 36; Conservative 32; Mismatches 62; Indels 23; Gaps 6;

Qy      8 QNGKEMSSSTTVSREDP-----ILPVYKGLEBK-----GYFDGWEISGP-EGKKDA 52
          : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
```

```
Db 576 EKGQEQAEDILIEEDFEFDLIGLQGLTDAEKAELKKAICAICYKPGATLFGGVNEGK--- 632
Qy 53 GVINLSKDTFTKPKVFKKIEEKEENKPTF-DVSKKKONQVNHSQLNESHKREDLQRE 111
Db 633 --LRSFROSLILSKVLKQOQDLLEEKNAVIODLAERVENYASLKKQCNFTIQDLLEIMVK 690
Qy 112 EHSQSDSTKQVDTATVLDKNNISSKSTTNNPNK 144
Db 691 BEHGALEKKDFVIQTM--EGSLAEVQVTENNKLK 721

RESULT 19
US-10-437-963-122282
; Sequence 122282, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122282
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_25224C.1.pep
US-10-437-963-122282

Query Match 12.4%; Score 93; DB 5; Length 869;
Best Local Similarity 45.1%; Pred. No. 16;
Matches 23; Conservative 7; Mismatches 17; Indels 4; Gaps 2;

Qy 84 DVSKKKONQVNHSQLNESHKREDLQREHSQSDSTKQVDTA--TVLDKNN 132
Db 617 DASKKDNHQSGNNL--SHRDEDTFRKKQKTNATSDACAQEVVTEKNN 665

RESULT 20
US-10-732-923-4286
; Sequence 4286, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 4286
; LENGTH: 948
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-732-923-4286

Query Match 12.4%; Score 93; DB 5; Length 948;
Best Local Similarity 21.4%; Pred. No. 18;
Matches 45; Conservative 29; Mismatches 56; Indels 80; Gaps 10;

Qy 1 HRVTVTQNGKMSITVSEEDFILPVY-----KGELEKGYQFDGWISGFEGKCD--- 51
Db 583 HYCTLSLQLEIKKDIQKFLNEETFLTKYGYYSFEKISLAINNSIDHY----FSHMKDNL 638
```

```
Qy 52 -----AGVINLSKDTFTKPKVFKKIEEKEENKPTF-----DVSK 87
Db 639 VICEPGRYMAASTLAVKIIGKR-----RPTFGIMLKDLKAHYDPLNFAQQENK 689
Qy 88 KCONPQVNHSQLNESHKRED-----LQREHSQK-----SDST----- 120
Db 690 KQDEPKINHNDNDNNNDNNNNNNNNNNNNQKGGQGNIMNDLIITNDSTNKKNDHSS 749
Qy 121 ----KQVDTATVLDK--NNIS-SKSTTNNPN 143
Db 750 SQVIONVSTIRDKEGDNKIKINTHTNNPN 779
```

## RESULT 21

```
US-10-732-923-8762
; Sequence 8762, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 8762
; LENGTH: 1529
; TYPE: PRT
; ORGANISM: Plasmodium yoelii yoelii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1529)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-8762
```

Query Match 12.4%; Score 93; DB 5; Length 1529;

Best Local Similarity 27.2%; Pred. No. 33; Indels 26; Gaps 5;  
Matches 28; Conservative 19; Mismatches 30;

```
Qy 57 NLSKDT-----FIKPFVKKIEEKEENKPTFDVSKKDNQVNHSQL-----NES 102
Db 356 NIISDTCIKIPIKVINSEYKNEEKEEKK-----NEKINDTIHYSESISKNSDNEQ 409
Qy 103 H-----RKEDLQREHSQKSDSTKQVDTATVLDKNN--ISSKSTT 139
Db 410 HPFLSKLRNVKKEEKKKKKKTKIKTVIAQKNKTVIAQKNKT 452
```

## RESULT 22

```
US-10-732-923-22588
; Sequence 22588, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 22588
; LENGTH: 3127
; TYPE: PRT
; ORGANISM: Plasmodium falciparum 3D7
US-10-732-923-22588
```

Query Match 12.4%; Score 93; DB 5; Length 3127;

Best Local Similarity 24.7%; Pred. No. 78;  
Matches 45; Conservative 29; Mismatches 58; Indels 50; Gaps 7;



```
; OTHER INFORMATION: accession number BL00412D, p-value=2.432e-11, raw score of 16.54
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1919)..(2122)
; OTHER INFORMATION: Neuraxin and MAP1B proteins domain identified by PFam,
; OTHER INFORMATION: accession name MAP1B_neuraxin, E-value=1.9e-59, PFam score of 190
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(2519)
; OTHER INFORMATION: Xaa = X or * as defined in Table 2
US-10-450-763-46995

Query Match          12.4%; Score 92.5; DB 5; Length 2519;
Best Local Similarity 31.0%; Pred. No. 66;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 58 LSKOTFIKPVFKIEKKEENKPTFDVSKKONPVNHSQNLNESHKEDLQRE-----E 112
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
689 VKKETKVP-----EDKKEKEKPKKEVAKKEDKTPI---KKEEKPKEEVKKEVKEIK 740
QY 113 HSKSDSTKOV 123
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
741 KEEKPEKKEV 751

RESULT 27
US-10-437-963-187665
; Sequence 187665, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 187665
; LENGTH: 1005
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_84346C.1.pep
US-10-437-963-187665

Query Match          12.3%; Score 92; DB 4; Length 1005;
Best Local Similarity 26.4%; Pred. No. 24;
Matches 32; Conservative 22; Mismatches 41; Indels 26; Gaps 4;

QY 45 GFEKKDAGYVNLKSDTFIKPVFKIEKKEENKPTFDVSKK-KONPV--NHSQNLN 101
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 GSEKEMSGKNIKSIKETGTGQSKELQKESKRSKTKDKSKKNKDMTQVPTNABEFHK 94
QY 102 SH-----RKE-----DLQREHSOKSDSTKQVTTATVLDKNNISSKST 138
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
95 EYTKVIRKESDSSIEQVIGTSSIQEMETNEQVSKDTSKDMTQVPANAGIRKEYT 154
QY 139 T 139
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
155 T 155

RESULT 28
US-10-282-122A-52328
; Sequence 52328, Application US/10282122A
; Publication No. US20040029129A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangou
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52328
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52328

Query Match          12.2%; Score 91.5; DB 4; Length 903;
Best Local Similarity 25.7%; Pred. No. 23;
Matches 38; Conservative 25; Mismatches 50; Indels 35; Gaps 8;

QY 5 VTIONGKEMSTIVSEEDFILPVYKGELEKGYQPDGHEISGPEGKKDAGYVNLKDTFI 64
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
190 INVELKEEKSKQIIEKHNEL-----KNBKSISPKESIKL---KKEKD----- 229
QY 65 KPVPKTEEKKEENKPTFDVSK-----KKDNPVNHSQNLNESHKEDLQREHSOKSD 118
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
230 KQVFAHNESNEENKQISKVNVELKNERSKQLPKIN-VELKEENKQSI--KEHNELRE 286
QY 119 STKQVTTATVLDKNNISSKSTNN--PNK 144
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
287 BTK----KCLPKVNIELKEETKKQVPNK 310

RESULT 29
US-10-381-596A-2
; Sequence 2, Application US/10381596A
; Publication No. US20040014178A1
; GENERAL INFORMATION:
; APPLICANT: Biostapro AB
; TITLE OF INVENTION: von Willebrand factor-binding proteins from
; TITLE OF INVENTION: Staphylococci
; FILE REFERENCE: 110059600
; CURRENT APPLICATION NUMBER: US/10/381,596A
```

;  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus lugdunensis  
 US-10-381-596A-2

## RESULT 31

APPLICANT: Griffais, R.

: NUMBER OF SEQ ID NOS: 6849

### Query Match

Query Match 12.0%; Score 89.5; DB 4; Length 511;  
Best Local Similarity 24.5%; Pred. No. 18;

Qy 64 IKPVTKIEKKKEENKPTFD-----VSKKDNPQVNHSQLNESHK 105

db 95 VKGVFKKTPOARPEVSSPRLEPSHVHVGORLPGLEGRDRIOKRSENPEADLGKMKRSYSD 154

.....

Qy 106 EDLQREHSQKSDSTKDVTA TVLDKNNISSKSTT 139

D6 155 GDLRVGHDSNEDSTEDSR-----EGGEPSSKSSS 185

D6 155 GDLRVGHDSNEDSTEDSR-----EGGEPSSKSSS 185

PERMIT 32

US-11-097-143-2904

; Sequence 2904, Application US/11097143

; Publication No. US20050208558A1

; GENERAL INFORMATION:

APPLICANT: Venter, J. Craig

APPLICANT: et al.

1. TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEAR

: TITLE OF INVENTION: DROSOPHILA GENES.

FILE REFERENCE: CL000728

; CURRENT APPLICATION NUMBER: US/11/097,143

**; CURRENT FILING DATE: 2005-04-04**

**; PRIOR APPLICATION NUMBER: 60/157,832**

; PRIOR FILING DATE: 1999-10-05

; PRIOR APPLICATION NUMBER: 60/160,191

; EXPIRATION DATE: 1000-10-10

; PRIOR FILING DATE: 1999-10-19  
 ; PRIOR APPLICATION NUMBER: 60/151 832

; PRIOR APPLICATION NUMBER: 60/161,332  
 : PRIOR FILING DATE: 1999-10-28

THE UNIVERSITY OF CHICAGO

```
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2904
; LENGTH: 1408
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-2904

Query Match      12.0%; Score 89.5; DB 6; Length 1408;
Best Local Similarity 24.7%; Pred. No. 60;
Matches 48; Conservative 23; Mismatches 62; Indels 61; Gaps 9;

Qy   9 NGKE-----MSSTIVSEEDP--ILPVYKGLEKGYPDGHWEISGFEGK--KDAGYV----- 55
     |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:  |||:
Db   646 NGKQKLNAKETKTTSKDFKEVIEYSKKE-EKNDNOESKELENDQDGKLPKDDQYLDVKPP 704

Qy   56 -----INLSKDTPIKVPFK----KIREKKEENKPTFDVSKKDNPNVNHSQLN 100
     :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:
Db   705 MESKEYSPLAVLKGEDVSXSPNESSDKTISEVKGE---PKMNGELPKTGEOVKQPEKS 762

Qy   101 ESHRK-EDLQREEHQSXSSTKDVDT-----ATVLDK 130
     ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:
Db   763 KTOAKIEFLAKVQSELKSKSVKDVTPKEQTTKETHTKGFSPKREATPKSEPKGTTEVDK 822

Qy   131 NNISSKSTNNPNK 144
     :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:  :||:
Db   823 SNDVSVIKPSEPDE 836


RESULT 33
US-10-437-963-160737
; Sequence 160737, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 160737
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_5998C.1.pep
US-10-437-963-160737

Query Match      11.9%; Score 89; DB 4; Length 519;
Best Local Similarity 26.0%; Pred. No. 20;
Matches 45; Conservative 18; Mismatches 44; Indels 66; Gaps 10;

Qy   5 VTIQNGKEM---SSSTIV-SEED-----FIL-----PVYKGBLEKG-YQF----- 38
     ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:  ::||:
Db   241 VVVEGGLMYRSGVLVNTTESKWTFVLTSTRSLRVGGOKKGKPOHSSFLAGAATTAAG 300
```

Db 470 NSAKKEATPATPSKTPSPVEKESQKDSQKDDNKQLPSVEKENDASSESGDKTPTATKP 529  
QY 129 DKNNISKSSTNNPNK 144  
Db 530 TKGEVSSSTT--PTK 543

## RESULT 35

US-10-470-048B-414  
; Sequence 414, Application US/10470048B  
; Publication No. US20050037444A1  
; GENERAL INFORMATION:  
; APPLICANT: MEINKE ET AL.  
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF  
; FILE REFERENCE: SONN:035US  
; CURRENT APPLICATION NUMBER: US/10/470,048B  
; CURRENT FILING DATE: 2003-07-25  
; NUMBER OF SEQ ID NOS: 603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 414  
; LENGTH: 645  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-470-048B-414

Query Match 11.9%; Score 89; DB 5; Length 645;  
Best Local Similarity 21.9%; Pred. No. 26;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;

QY 6 TIQNGKMSSTIVSEEDFILPVYKGELE-KGYQF-----DGWEISGPEGKK----- 50  
Db 355 SVENNESMMDTFVKH-----PIKTGMLNGKKYVMVMTTNDYDKDFMVEGQRTISKDA 409  
QY 51 -----DAGVIVNL-SKDTFIKPVFKIIEKKKEE 78  
Db 410 KNTRTIIPPVVEGKTLDAIVKVVHTIDYDQGHVVRIVDKAFTKANTDKSNKKEOOD 469  
QY 79 NKPTFDV-----SKKONPQVNHSQLNESHKREDLQ-----RREHSQKSDSTKDVT-ATVL 128  
Db 470 NSAKKEATPATPSKTPSPVEKESQKDSQKDDNKQLPSVEKENDASSESGDKTPTATKP 529

QY 129 DKNNISKSSTNNPNK 144

Db 530 TKGEVSSSTT--PTK 543

## RESULT 36

US-10-172-502-10  
; Sequence 10, Application US/10172502  
; Publication No. US20030185833A1  
; GENERAL INFORMATION:  
; APPLICANT: FOSTER, Timothy et al.  
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES...  
; FILE REFERENCE: P07263US01/BAS  
; CURRENT APPLICATION NUMBER: US/10/172,502  
; CURRENT FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: US 60/298,098  
; PRIOR FILING DATE: 2001-06-15  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 654  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-172-502-10

Query Match 11.9%; Score 89; DB 4; Length 654;  
Best Local Similarity 21.9%; Pred. No. 26;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;

QY 6 TIQNGKMSSTIVSEEDFILPVYKGELE-KGYQF-----DGWEISGPEGKK----- 50

Db 364 SVENNESMMDTFVKH-----PIKTGMLNGKKYVMVMTTNDYDKDFMVEGQRTISKDA 418  
QY 51 -----DAGVIVNL-SKDTFIKPVFKIIEKKKEE 78  
Db 419 KNTRTIIPPVVEGKTLDAIVKVVHTIDYDQGHVVRIVDKAFTKANTDKSNKKEOOD 478  
QY 79 NKPTFDV-----SKKONPQVNHSQLNESHKREDLQ-----RREHSQKSDSTKDVT-ATVL 128  
Db 479 NSAKKEATPATPSKTPSPVEKESQKDSQKDDNKQLPSVEKENDASSESGDKTPTATKP 538  
QY 129 DKNNISKSSTNNPNK 144  
Db 539 TKGEVSSSTT--PTK 552

## RESULT 37

US-11-020-509-10  
; Sequence 10, Application US/11020509  
; Publication No. US20050106648A1  
; GENERAL INFORMATION:  
; APPLICANT: FOSTER, Timothy et al.  
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES...  
; FILE REFERENCE: P07263US02/BAS  
; CURRENT APPLICATION NUMBER: US/11/020,509  
; CURRENT FILING DATE: 2004-12-27  
; PRIOR APPLICATION NUMBER: US 10/172,502  
; PRIOR FILING DATE: 2002-06-17  
; PRIOR APPLICATION NUMBER: US 60/298,098  
; PRIOR FILING DATE: 2001-06-15  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 654  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-11-020-509-10

Query Match 11.9%; Score 89; DB 6; Length 654;  
Best Local Similarity 21.9%; Pred. No. 26;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;

QY 6 TIQNGKMSSTIVSEEDFILPVYKGELE-KGYQF-----DGWEISGPEGKK----- 50  
Db 364 SVENNESMMDTFVKH-----PIKTGMLNGKKYVMVMTTNDYDKDFMVEGQRTISKDA 418  
QY 51 -----DAGVIVNL-SKDTFIKPVFKIIEKKKEE 78  
Db 419 KNTRTIIPPVVEGKTLDAIVKVVHTIDYDQGHVVRIVDKAFTKANTDKSNKKEOOD 478  
QY 79 NKPTFDV-----SKKONPQVNHSQLNESHKREDLQ-----RREHSQKSDSTKDVT-ATVL 128  
Db 479 NSAKKEATPATPSKTPSPVEKESQKDSQKDDNKQLPSVEKENDASSESGDKTPTATKP 538  
QY 129 DKNNISKSSTNNPNK 144  
Db 539 TKGEVSSSTT--PTK 552

## RESULT 38

US-11-097-143-21276  
; Sequence 21276, Application US/11097143  
; Publication No. US20050208558A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; TITLE OF INVENTION: DROSOPHILA GENES.  
; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2006, 15:00:02 ; Search time 22.3893 Seconds  
(without alignments)  
531.741 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773  
Perfect score: 748  
Sequence: 1 HRVTVTQNGKEMSTIVSE.....ATVLDKNNISSKSTNNPNK 144

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgm2\_6/ptodata/1/iaa/5\_COMB.pep:\*  
2: /cgm2\_6/ptodata/1/iaa/6\_COMB.pep:\*  
3: /cgm2\_6/ptodata/1/iaa/H\_COMB.pep:\*  
4: /cgm2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgm2\_6/ptodata/1/iaa/RE\_COMB.pep:\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	748	100.0	773	2	US-09-590-991-8
2	745	99.6	637	2	US-09-107-433-3169
3	745	99.6	2138	2	US-09-583-110-5274
4	615	82.2	117	2	US-08-961-083-68
5	615	82.2	117	2	US-09-536-784-68
6	615	82.2	117	2	US-09-765-271-68
7	615	82.2	117	2	US-09-765-272A-68
8	110	14.7	746	2	US-09-710-279-652
9	110	14.7	778	2	US-09-134-001C-3868
10	101.5	13.6	347	2	US-09-248-796A-16224
11	95	12.7	348	2	US-09-538-092-1316
12	92.5	12.4	2468	2	US-09-976-594-726
13	92.5	12.4	2468	2	US-09-538-092-1135
14	92.5	12.4	2522	2	US-09-949-016-10237
15	89.5	12.0	511	2	US-09-198-452A-509
16	89.5	12.0	511	2	US-09-438-185A-475
17	89	11.9	654	2	US-10-172-502-10
18	88	11.8	280	2	US-09-248-796A-17646
19	87.5	11.7	1702	2	US-08-296-791-5
20	87.5	11.7	1702	2	US-09-839-986-5
21	87.5	11.7	1702	2	US-10-080-505-5
22	87.5	11.7	1702	2	US-10-645-655-5
23	87.5	11.7	1702	4	PCT-US95-10661A-5
24	87	11.6	902	2	US-09-134-001C-5157
25	86	11.5	172	2	US-09-248-796A-21065
26	86	11.5	243	2	US-09-248-796A-20306
27	86	11.5	700	2	US-08-235-836C-74

28	86	11.5	758	2	US-09-949-016-8288	Sequence 8288, Ap
29	86	11.5	1315	2	US-09-200-850B-5	Sequence 5, Appli
30	86	11.5	1989	2	US-09-949-016-10076	Sequence 10076, A
31	85.5	11.4	109	2	US-09-248-796A-24668	Sequence 24668, A
32	85.5	11.4	402	2	US-09-464-483-4	Sequence 4, Appli
33	85.5	11.4	402	2	US-09-414-664-4	Sequence 4, Appli
34	85.5	11.4	529	2	US-09-464-483-2	Sequence 2, Appli
35	85.5	11.4	529	2	US-09-414-664-2	Sequence 2, Appli
36	85	11.4	465	2	US-09-134-001C-3856	Sequence 3856, App
37	85	11.4	472	2	US-09-710-279-658	Sequence 658, App
38	85	11.4	700	1	US-07-720-589-2	Sequence 2, Appli
39	85	11.4	700	1	US-08-785-190-2	Sequence 2, Appli
40	85	11.4	700	2	US-08-235-836C-66	Sequence 66, Appl
41	85	11.4	700	4	PCT-US92-05539-2	Sequence 2, Appli
42	85	11.4	708	2	US-08-235-836C-76	Sequence 76, Appl
43	85	11.4	720	2	US-09-710-279-2058	Sequence 2058, Ap
44	85	11.4	728	2	US-09-134-001C-4968	Sequence 4968, Ap
45	85	11.4	743	2	US-08-910-925-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-590-991-8  
; Sequence 8, Application US/09590991  
; Patent No. 6887480  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; APPLICANT: Choi, Gil  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-475  
; CURRENT APPLICATION NUMBER: US/09/590,991  
; CURRENT FILING DATE: 2000-06-09  
; EARLIER APPLICATION NUMBER: U.S. 60/138,453  
; EARLIER FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-590-991-8

Query Match	100.0%;	Score 748;	DB 2;	Length 773;
Best Local Similarity	100.0%;	Pred. No. 5.9e-71;		
Matches 144;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	HRVTVTQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGYVINLSK	60	
Db	630	HRVTVTQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFGKKDAGYVINLSK	689	
Qy	61	DTFKPVFKLIEKKKEBKPTDVSKKDNPQVNHSQLNESHKEDLQREHHSOKSDST	120	
Db	690	DTFKPVFKLIEKKKEBKPTDVSKKDNPQVNHSQLNESHKEDLQREHHSOKSDST	749	
Qy	121	KDVTATVLDKNNISSKSTNNPNK	144	
Db	750	KDVTATVLDKNNISSKSTNNPNK	773	

RESULT 2

US-09-107-433-3169  
; Sequence 3169, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```
/ STREET: 100 Beaver Street
/ CITY: Waltham
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02354
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: CD/ROM ISO9660
/ COMPUTER: <Unknown>
/ OPERATING SYSTEM: <Unknown>
/ SOFTWARE: <Unknown>
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/107,433
/ FILING DATE: 30-Jun-1998
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/ 085131
/ FILING DATE: May 12, 1998
/ APPLICATION NUMBER: 60/051553
/ FILING DATE: July 2, 1997
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-011
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/
/ INFORMATION FOR SEQ ID NO: 3169:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 637 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Streptococcus pneumoniae
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (B) LOCATION 1...637
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
/
/ US-09-107-433-3169
/
/ Query Match 99.6%; Score 745; DB 2; Length 637;
/ Best Local Similarity 99.3%; Pred. No. 9.5e-71;
/ Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 60
/ Db 460 HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 519
/
/ QY 61 DTFIKPVFKIEEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHHSQKSDST 120
/ Db 520 DTFIKPVFKIEEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHHSQKSDST 579
/
/ QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
/ Db 580 KDVTATVLDKNNISSKSTTNNPNK 603
/
/ RESULT 3
/ US-09-583-110-5274
/ Sequence 5274, Application US/09583110
/ Patent No. 6699703
/
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al.
/ TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
/ FILE REFERENCE: PATH00-07A
/ CURRENT APPLICATION NUMBER: US/09/583,110
/ CURRENT FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/107,433
/ PRIOR FILING DATE: 1998-06-30
/ PRIOR APPLICATION NUMBER: US 60/085,131
/ PRIOR FILING DATE: 1998-05-12
/ PRIOR APPLICATION NUMBER: US 60/051,553
/
```

```
/ PRIOR FILING DATE: 1997-07-02
/ NUMBER OF SEQ ID NOS: 5322
/ SEQ ID NO 5274
/ LENGTH: 2138
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
/ US-09-583-110-5274
/
/ Query Match 99.6%; Score 745; DB 2; Length 2138;
/ Best Local Similarity 99.3%; Pred. No. 4.9e-70;
/ Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 60
/ Db 1961 HRVTVIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSK 2020
/
/ QY 61 DTFIKPVFKIEEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHHSQKSDST 120
/ Db 2021 DTFIKPVFKIEEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHHSQKSDST 2080
/
/ QY 121 KDVTATVLDKNNISSKSTTNNPNK 144
/ Db 2081 KDVTATVLDKNNISSKSTTNNPNK 2104
/
/ RESULT 4
/ US-08-961-083-68
/ Sequence 68, Application US/08961083
/ Patent No. 6159469
/ GENERAL INFORMATION:
/ APPLICANT: Choi et. al.
/ TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
/ NUMBER OF SEQUENCES: 452
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII Text
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/961,083
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Brookes, A. Anders
/ REGISTRATION NUMBER: 36,373
/ REFERENCE/DOCKET NUMBER: PB340P2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (301) 309-8504
/ TELEFAX: (301) 309-8512
/ INFORMATION FOR SEQ ID NO: 68:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 117 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-961-083-68
/
/ Query Match 82.2%; Score 615; DB 2; Length 117;
/ Best Local Similarity 100.0%; Pred. No. 6.7e-58;
/ Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 28 YKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTFIKPVFKIEEKKKEENKPTFDVSK 87
```

Db 1 YGGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDFIKPVFKKIEKKEEENKPTFDVSK 60  
QY 88 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 144  
Db 61 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117

## RESULT 5

US-09-536-784-68  
; Sequence 68, Application US/09536784  
; Patent No. 6573082

## GENERAL INFORMATION:

APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/536,784  
FILING DATE: 30-Oct-1997  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083  
FILING DATE: OCT-30-1997  
ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB340P3

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

## INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 68:

US-09-536-784-68

Query Match 82.2%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 6.7e-58;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 YGGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDFIKPVFKKIEKKEEENKPTFDVSK 87

Db 1 YGGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDFIKPVFKKIEKKEEENKPTFDVSK 60

QY 88 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 144

Db 61 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117

## RESULT 6

US-09-765-271-68  
; Sequence 68, Application US/09765271  
; Patent No. 6887663

## GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,271  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/536,784  
FILING DATE: <Unknown>  
APPLICATION NUMBER: 08/961,083  
FILING DATE: OCT-30-1997

## ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks  
REGISTRATION NUMBER: 41,971  
REFERENCE/DOCKET NUMBER: PB340P3

## TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

## INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 68:

US-09-765-271-68

Query Match 82.2%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 6.7e-58;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 YGGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDFIKPVFKKIEKKEEENKPTFDVSK 87

Db 1 YGGELEKGYQFDGWEISGFEKGDAGYVNLSDTKDFIKPVFKKIEKKEEENKPTFDVSK 60

QY 88 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 144

Db 61 KKDNPQVNHSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 117

## RESULT 7

US-09-765-272A-68  
; Sequence 68, Application US/09765272A  
; Patent No. 6929930

## GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and

NUMBER OF SEQUENCES: 454

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: Dell Latitude C610  
OPERATING SYSTEM: Windows 2000  
SOFTWARE: ASCII Text

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272A  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: OCT-30-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Lin J. Hymel  
REGISTRATION NUMBER: 45,414  
REFERENCE/DOCKET NUMBER: PB340P2C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 610-5790  
TELEFAX: (301) 309-8439  
INFORMATION FOR SEQ ID NO: 68:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 68:  
US-09-765-272A-68

Query Match 82.2%; Score 615; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 6.7e-58;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 YGELEKGYQPGWISGPEGKKGAGYVNLKOTFIKPVFKIIEKKEENKPTFDVSK 87

DB 1 YGELEKGYQPGWISGPEGKKGAGYVNLKOTFIKPVFKIIEKKEENKPTFDVSK 60

QY 88 KKDNPQVNHSQLNESHKEDLQREHSHOKSDSTKDVATVLDKNNISSKSTNNPK 144

DB 61 KKDNPQVNHSQLNESHKEDLQREHSHOKSDSTKDVATVLDKNNISSKSTNNPK 117

## RESULT 8

US-09-710-279-652

Sequence 652, Application US/09710279

Patent No. 6703492

GENERAL INFORMATION:

APPLICANT: KIMMERLY, WILLIAM JOHN

TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

FILE REFERENCE: PU3480US

CURRENT APPLICATION NUMBER: US/09/710,279

CURRENT FILING DATE: 2000-11-09

PRIOR APPLICATION NUMBER: 60/164,258

PRIOR FILING DATE: 1999-11-09

NUMBER OF SEQ ID NOS: 4472

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 652

LENGTH: 746

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic

OTHER INFORMATION: amino acid sequence

US-09-710-279-652

Query Match 14.7%; Score 110; DB 2; Length 746;  
Best Local Similarity 27.0%; Pred. No. 0.0048;  
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;

QY 5 VTIOGKEMSSTIVSEEDFILPVYK-----GELEKGYQFDGW---EISGPE-----G 48

DB 596 ITIGNGQIKQKQSVKSGTKVLPHSKVLMWDGELTMP-DMTGWTKEDVLAFLADLTKLKVS 654

QY 49 KKDAGYVNL--LSKDTFIKPVFKIIEKKEENKPTFDVS-----KKDNPQVNHSQLNES 102

DB 655 TKGNGFVTNQSIKSGQIIK-----NKKIEVLSABTDQDQKTDDESDSN 701

QY 103 HRKEDLQREHSHOKSDSTKDVATVLDKNNISSKSTNN 141

DB 702 KSKDKAKADEHSHNTSSSTKN-----DKSNADSKNDSDD 734

## RESULT 9

US-09-134-001C-3868

Sequence 3868, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: LYNN Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 3868

LENGTH: 778

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3868

Query Match 14.7%; Score 110; DB 2; Length 778;

Best Local Similarity 27.0%; Pred. No. 0.0051;

Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;

QY 5 VTIOGKEMSSTIVSEEDFILPVYK-----GELEKGYQFDGW---EISGPE-----G 48

DB 628 ITIGNGQIKQKQSVKSGTKVLPHSKVLMWDGELTMP-DMTGWTKEDVLAFLADLTKLKVS 686

QY 49 KKDAGYVNL--LSKDTFIKPVFKIIEKKEENKPTFDVS-----KKDNPQVNHSQLNES 102

DB 687 TKGNGFVTNQSIKSGQIIK-----NKKIEVLSABTDQDQKTDDESDSN 733

QY 103 HRKEDLQREHSHOKSDSTKDVATVLDKNNISSKSTNN 141

DB 734 KSKDKAKADEHSHNTSSSTKN-----DKSNADSKNDSDD 766

## RESULT 10

US-09-248-796A-16224

Sequence 16224, Application US/09248756A

Patent No. 6747137

GENERAL INFORMATION:

APPLICANT: Keith Weinstein et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.132

CURRENT APPLICATION NUMBER: US/09/248,796A

CURRENT FILING DATE: 1999-02-12

PRIOR APPLICATION NUMBER: US 60/074,725

PRIOR FILING DATE: 1998-02-13

PRIOR APPLICATION NUMBER: US 60/096,409

PRIOR FILING DATE: 1998-08-13

NUMBER OF SEQ ID NOS: 28208

SEQ ID NO 16224

LENGTH: 347

TYPE: PRT

ORGANISM: Candida albicans

US-09-248-796A-16224

Query Match 13.6%; Score 101.5; DB 2; Length 347;

Best Local Similarity 21.4%; Pred. No. 0.014;

Matches 31; Conservative 32; Mismatches 63; Indels 19; Gaps 4;

QY 12 EMSSTIVSEEDFILPVYKGELEKGYQFDGWISGPF-----EGKKDAGYVNLKOTFIK 65

DB 202 KLSPLMDEIDTPEIYND--KKWYVATSLQVYQTDLESSESGIGWEDDLEENYRTG 259

QY 66 FVPFKIIEKKEENKPTFDVSKKDNQVNHSQLN-----ESHRKEDLQREHSHOKSDS 119

Db 260 PVFKTLQDLRE-----WRAEKBOANPKKBEENLNQKPVAKQKQKQPNSTKQKQKQKQKQ 314  
Qy 120 TKDVTATVLDKNNISKSSTNNPNK 144  
Db 315 TKKITPKTSKMLSGISTNLIINK 339

RESULT 11  
US-09-538-092-1316  
; Sequence 1316, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 1316  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number Q14093  
US-09-538-092-1316

Query Match 12.7%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 0.068;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;  
Qy 29 KGELEKGFQDGEISGEGKKDAGVINLSKDTPIKPVKKIEBKKEENKPTP---DV 85  
Db 205 ESEGEKG---GTEKDSKKGKDS-----KKGKDSAIQLQAVKADKKDDEKGDANKGDE 256  
Qy 86 SK--KKDNPQVNHSQLN-----ESHKEDLQREHSQKSDSTKD---VTATVLDKNNI 133  
Db 257 SKDAKKDAKEIKKGKKKKKPSSTDSKDDVKKB---SKDATKQAKKVAKKDTEKESA 313  
Qy 134 SSK 136  
Db 314 DSK 316

RESULT 12  
US-09-976-594-726  
; Sequence 726, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 726  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature

; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1  
US-09-976-594-726

Query Match 12.4%; Score 92.5; DB 2; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 1.8;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
Qy 58 LSKDTFIKPVFKJIEBKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQRE-----E 112  
Db 638 VKKETKVKP-----EDKKEKEKPKKEVAKEDKTPI---KKEBKPKKEEVKKEVKKEIK 689  
Qy 113 HSQKSDSTKDV 123  
Db 690 KEEKKPKKEV 700

RESULT 13  
US-09-538-092-1135  
; Sequence 1135, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CurapatSeqFormatter Version 0.9  
; SEQ ID NO 1135  
; LENGTH: 2468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P46821  
US-09-538-092-1135

Query Match 12.4%; Score 92.5; DB 2; Length 2468;  
Best Local Similarity 31.0%; Pred. No. 1.8;  
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;  
Qy 58 LSKDTFIKPVFKJIEBKKEENKPTFDVSKKQNPQVNHSQLNESHKEDLQRE-----E 112  
Db 638 VKKETKVKP-----EDKKEKEKPKKEVAKEDKTPI---KKEBKPKKEEVKKEVKKEIK 689  
Qy 113 HSQKSDSTKDV 123  
Db 690 KEEKKPKKEV 700

RESULT 14  
US-09-949-016-10237  
; Sequence 10237, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USBS THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08





```
; TITLE OF INVENTION: HAEMOPHILUS ADHERENCE AND PENETRATION PROTEINS
; FILE REFERENCE: A-59941-1/RFT/DCF/DHR
; CURRENT APPLICATION NUMBER: US/10/080,505
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: US 08/296,791
; PRIOR FILING DATE: 1994-10-25
; PRIOR APPLICATION NUMBER: US 09/839,996
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-10-080-505-5

Query Match      11.7%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 3.6;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY      56 INLSKDTFKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREEHS- 114
      ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY      115 --QKSDSTKDVATVTLDKNNISSKSTNNPNK 144
      : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 22
US-10-645-655-5
; Sequence 5, Application US/10645655
; Patent No. 6815182
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/645,655
; FILING DATE: 20-Aug-2003
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
; US-10-645-655-5
```

```
Query Match      11.7%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred. No. 3.6;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY      56 INLSKDTFKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREEHS- 114
      ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY      115 --QKSDSTKDVATVTLDKNNISSKSTNNPNK 144
      : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 23
PCT-US95-10661A-5
; Sequence 5, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: EP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; PCT-US95-10661A-5

Query Match      11.7%; Score 87.5; DB 4; Length 1702;
Best Local Similarity 27.2%; Pred. No. 3.6;
Matches 25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

QY      56 INLSKDTFKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREEHS- 114
      ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1296 INTGSATAITETAESKDKPQTETAASTEDASQHKANTVADNSVANNSESSEPKSRRRSI 1355

QY      115 --QKSDSTKDVATVTLDKNNISSKSTNNPNK 144
      : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1356 SQPQTSABETTAASTDETTIADNSKRSKPNR 1387

RESULT 24
US-09-134-001C-5157
; Sequence 5157, Application US/09134001C
; Patent No. 6380370
```



```

RESULT 25
US-09-248-796A-21065
; Sequence 21065, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 21065
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-21065

Query Match 11.5%; Score 86; DB 2; Length 172;
Best Local Similarity 22.1%; Pred. No. 0.24;
Matches 34; Conservative 25; Mismatches 67; Indels 28; Gaps 4;

Qy 1 HRVTVT--IQNGKEMSSTIVSEED-----FILPVYKGEL-----EKGYQFDGWEISGF 46
Db 9 HPVVKTRWVGSKELISQGFEEEDGTPKPYFFKYNVQLAIPRPVDEDTYDLWIEIKY 68
Qy 47 EGKQDAGVYNLSKDTPIKPVFKKIIEKKGEEENKPTPDVSKCKDNPNQVNSQLNESHRKE 106
Db 69 ESK-----MKERARREKEREREKRDLEKKKKQKQKQKQKQKQKQKQKQKQKQKQK 114
Qy 107 DLQREHSQKSDSTKQVTAIVLDKNNISSKSTTN 140
Db 115 DDEKNQDTRNNTDKKDSEKQKSDKPTVEAKKETD 148

```

RESULT 27  
US-08-235-836C-74  
; Sequence 74, Application US/08235836C  
; Patent No. 6248562  
; GENERAL INFORMATION:  
; APPLICANT: Dunn, John J.  
; APPLICANT: Luft, Benjamin J.  
; TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising  
; TITLE OF INVENTION: Borrelia Poly peptides and Uses Therefor  
; NUMBER OF SEQUENCES: 144  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brookhaven National Laboratory  
; STREET:  
; CITY: Upton  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 11973  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/235,836C  
; FILING DATE: 29-APR-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/148,191  
; FILING DATE: 01-11-93  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bogosian, Margaret C.  
; REGISTRATION NUMBER: 25,324  
; REFERENCE/DOCKET NUMBER: BNL93-28A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516) 282-7338  
; TELEFAX: (516) 282-3729  
; INFORMATION FOR SEQ ID NO: 74:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 700 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULAR TYPE: protein  
US-08-235-836C-74

Query Match 11.5%; Score 86; DB 2; Length 700;  
Best Local Similarity 23.2%; Pred. No. 1.6;  
Matches 41; Conservative 38; Mismatches 54; Indels 44; Gaps 9;  
QY 3 VVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFD-----GWEISGPEGKK 50  
Db ITETIENLRDLEKATDEE-----HKKEIES--QVDAKKQKBEKLDKKAIDLDKAQQKL 309  
QY 51 D-AGVINLSKDTFKPKVPFKIEKKEENKPTFDVSKKONPQVNH-----HSQLN 100  
Db 310 DPAEDNLDIQDVTREKLQENINETNKEKILPKPGDVSSPKYDKQLQIKESLEDLQEQLK 369  
QY 101 ES--HRKEDLQREHSQKSD-----STKDVATVLDKNNISSKSTNNPNK 144  
Db 370 ETGDENQREIEKQIEIKKSDKLLKSKDDKASKDGKALDLDR-ELNSKASSKEKSK 425

RESULT 28  
US-09-949-016-8288  
; Sequence 8288, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8288  
; LENGTH: 758  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-8288

Query Match 11.5%; Score 86; DB 2; Length 758;  
Best Local Similarity 29.1%; Pred. No. 1.8;  
Matches 44; Conservative 13; Mismatches 52; Indels 42; Gaps 8;  
QY 9 NGKMSSTIVSE--BDFILPVYKGELEKGYQF-----DGWE-----ISGFEKK 50  
Db 410 HGKSDSPNVYTEKKIARLRLTELKLTPEQQRSDLWERYLYEAKDQNGKQGTGDK 469  
QY 51 DAG-----YVINLSKDTFKPKVPFKIEKKEENKPTFDVSKKONPQVNH-----SQLNE 101  
Db 470 KGGRGSHRAKNKSKETFLGSV-----KETFDAMKNSTKEFVRHHKEKIKQAKE 517  
QY 102 SHRKEDLQREHSQKSD--STKDVATVLDK 130  
Db 518 A-VKENLKKFSDSVKSTFRHFKDITKNIFDE 547

RESULT 29  
US-09-200-650E-5  
; Sequence 5, Application US/09200650E  
; Patent No. 6680195  
; GENERAL INFORMATION:  
; APPLICANT: Patti, Joseph M.  
; APPLICANT: Foster, Timothy J.

; APPLICANT: Hook, Magnus A.O.  
; APPLICANT: Eichinn, Deirdre Ni  
; APPLICANT: Perkins, Samuel L.  
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus  
; FILE REFERENCE: P06283US2/BAS  
; CURRENT APPLICATION NUMBER: US/09/200,650E  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,815  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: 60/098,427  
; PRIOR FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1315  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-200-650E-5

Query Match 11.5%; Score 86; DB 2; Length 1315;  
Best Local Similarity 24.2%; Pred. No. 3.7;  
Matches 36; Conservative 22; Mismatches 65; Indels 26; Gaps 4;  
QY 3 VVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAGVINLSK 60  
Db 819 VVTILKN-----ENGEVLQTTTKTDKDKGYQFTGLENGTYKVEFETSGYTPT--- 865  
QY 61 DTFKPKVPFKIEKKEENKPTFDVSKKONPQVNHSQLNESHKEDLQREHSQKSDST 120  
Db 866 -----QVGSQTDGIDSGNSTTGTGVIKDKONDTIDSGFYKPTYNLGDYVWEDTNKNGVQD 920  
QY 121 KD-----VTATVLDKNNISSKSTNNPN 143  
Db 921 KDEIGISGVTVLKDENDKVLKIVTTDEN 949

RESULT 30  
US-09-949-016-10076  
; Sequence 10076, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10076  
; LENGTH: 1989  
; TYPE: PRT  
; ORGANISM: Human  
US-09-949-016-10076

Query Match 11.5%; Score 86; DB 2; Length 1989;  
Best Local Similarity 26.2%; Pred. No. 6.5;  
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;  
QY 37 QFDGWEISGFEGKKDAGVINLSKDTFKPKVPFKIEKKEENKPTFDVSKKONPQVNH 96  
Db 1002 EMNNLIQISVIRIKKGVAVT--KLKVHAFMQAHFK---QREADEVKPLDELYEKKANCIAIH 1057  
QY 97 SQLNESHKEDLQREHSQKSDSTKUTATVLDKNNISSKSTNNPN 143  
Db 1058 TGA-DIHRNGDFQKNGTGTSGIGSVYEKVIIDEDHM---SFINPN 1100

RESULT 32  
US-09-464-483--4  
; Sequence 4, Application US/09464483  
; Patent No. 6228617  
; GENERAL INFORMATION:  
; APPLICANT: Lawlor, Elizabeth J.  
; TITLE OF INVENTION: No. 6228617el tig  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Street  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/464,483  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/999,339  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Falk, Stephen T  
; REGISTRATION NUMBER: 36,795  
; REFERENCE/DOCKET NUMBER: GM10085  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2488  
; TELEFAX: 215-994-2222  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:

```

1  RESULT 33
2  US-09-414-664-4
3  ; Sequence 4, Application US/09414664
4  ; Patent No. 6242249
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Lawlor, Elizabeth J.
7  ; TITLE OF INVENTION: No. 6242249el tlg
8  ; NUMBER OF SEQUENCES: 6
9  ; CORRESPONDENCE ADDRESS:
10 ; ADDRESSEE: Dechert Price & Rhoads
11 ; STREET: 4000 Bell Atlantic Tower, 1717 Arch Street
12 ; CITY: Philadelphia
13 ; STATE: PA
14 ; COUNTRY: US
15 ; ZIP: 19103
16 ; COMPUTER READABLE FORM:
17 ; MEDIUM TYPE: Diskette
18 ; COMPUTER: IBM Compatible
19 ; OPERATING SYSTEM: DOS
20 ; SOFTWARE: FastSeq for Windows Version 2.0
21 ; CURRENT APPLICATION DATA:
22 ; APPLICATION NUMBER: US/09/414,664
23 ; FILING DATE:
24 ; CLASSIFICATION:
25 ; PRIOR APPLICATION DATA:
26 ; APPLICATION NUMBER: 08/999,339
27 ; FILING DATE:
28 ; ATTORNEY/AGENT INFORMATION:
29 ; NAME: Falk, Stephen T.
30 ; REGISTRATION NUMBER: 36,795
31 ; REFERENCE/DOCKET NUMBER: GMI0085
32 ; TELECOMMUNICATION INFORMATION:
33 ; TELEPHONE: 215-994-2488
34 ; TELEFAX: 215-994-2222
35 ; TELEX:
36 ; INFORMATION FOR SEQ ID NO: 4:
37 ; SEQUENCE CHARACTERISTICS:
38 ; LENGTH: 402 amino acids
39 ; TYPE: amino acid
40 ; STRANDEDNESS: single
41 ; TOPOLOGY: linear
42 ; US-09-414-664-4

```

Query Match 11.4%; Score 85.5; DB 2; Length 402;  
Best Local Similarity 22.9%; Pred. No. 0.84;  
Matches 49; Conservative 30; Mismatches 52; Indels 83; Gaps 13;





```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/785,190
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/720,589
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 2307U-330/UC 90-017-1
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-785-190-2

Query Match 11.4%; Score 85; DB 1; Length 700;
Best Local Similarity 24.4%; Pred. No. 2;
Matches 44; Conservative 33; Mismatches 53; Indels 50; Gaps 9;

Qy 3 VVTVIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWEISGREGKDGAGVYNLSK-- 60
Db 258 ITETIENLRDQLEKATDEE-----HKKEIES--QVDAKKQKEELDKA---INLDKRAQ 306
Qy 61 -----DTPIKPVFKIEKKKEENKP-TFDVSKKKDNQPQN-----HS 97
Db 307 QKLDSAEADNLDVQRNTVREKIQEDINEINKEKNLPKPGDVSSPKVDKQLQIKESLEDLQE 366
Qy 98 QLNES---HRKEDLQREHSQKSD-----STKDVATATVLDKNNISSKSTNNPNK 144
Db 367 QLKETGDNQKREIEKQIEIKSDKLLKSKDKKDKGKALDLDLR-ELNSKASSKEKSK 425

RESULT 40
US-08-235-836C-66
; Sequence 66, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Brookhaven National Laboratory
; STREET:
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,836C
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93
```

```
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Bogosian, Margaret C.
; REGISTRATION NUMBER: 25,324
; REFERENCE/DOCKET NUMBER: BNL93-28A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 282-7338
; TELEFAX: (516) 282-3729
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-235-836C-66

Query Match 11.4%; Score 85; DB 2; Length 700;
Best Local Similarity 24.4%; Pred. No. 2;
Matches 44; Conservative 33; Mismatches 53; Indels 50; Gaps 9;

Qy 3 VVTVIQNGKMSSTIVSEEDFLPVYKGELEKGYQFDGWEISGREGKDGAGVYNLSK-- 60
Db 258 ITETIENLRDQLEKATDEE-----HKKEIES--QVDAKKQKEELDKA---INLDKRAQ 306
Qy 61 -----DTPIKPVFKIEKKKEENKP-TFDVSKKKDNQPQN-----HS 97
Db 307 QKLDSAEADNLDVQRNTVREKIQEDINEINKEKNLPKPGDVSSPKVDKQLQIKESLEDLQE 366
Qy 98 QLNES---HRKEDLQREHSQKSD-----STKDVATATVLDKNNISSKSTNNPNK 144
Db 367 QLKETGDNQKREIEKQIEIKSDKLLKSKDKKDKGKALDLDLR-ELNSKASSKEKSK 425

Search completed: April 24, 2006, 15:03:37
Job time : 23.3893 secs
```